

SEARCH REQUEST FORM

Requestor's Name: _____ Serial Number: _____
Date: _____ Phone: _____ Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

Shears, Beverly

From: Devi, Sarvamangala
Sent: Wednesday, January 28, 2004 9:02 AM
To: Shears, Beverly
Subject: 09/359,426

Beverly:

Would you please perform a sequence and an interference search on SEQ ID NO: 2 in application SN 09/359,426?

Thanks.

S. DEVI, Ph.D.
AU 1645
Remsen - 3B07



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 112984

To: Sarvamangala Devi
Location: REM 3B07
Art Unit: 1645
Friday, January 30, 2004

Case Serial Number: 09/359426

From: Beverly Shears
Location: Remsen Bldg.
RM 1A54
Phone: 571-272-2528

beverly.shears@uspto.gov

STAFF USE ONLY

Date completed: 01-29-04
Searcher: Beverly 2528
Terminal time: 22
Elapsed time: _____
CPU time: _____
Total time: 25
Number of Searches: _____
Number of Databases: 1

Search Site

____ STIC
____ CM-1
____ Pre-S

Type of Search

____ N.A. Sequence
____ A.A. Sequence
____ Structure
____ Bibliographic

Vendors

____ IG
____ STN
____ Dialog
____ APS
____ Geninfo
____ SDC
____ DARC/Questel
____ Other CGN

THIS PAGE BLANK (USPTO)

Devi, S.
09/359426

09/359426

(FILE 'REGISTRY' ENTERED AT 15:43:31 ON 29 JAN 2004)
L2 1 S EEKTPLTAA.APVV.NA/SQSP

FILE 'HCAPLUS' ENTERED AT 15:44:32 ON 29 JAN 2004
L3 1 S L2

L3 ANSWER 1 OF 1 HCAPLUS COPYRIGHT 2004 ACS on STN
ACCESSION NUMBER: 1998:527345 HCAPLUS
DOCUMENT NUMBER: 129:160619
TITLE: Pseudomonas aeruginosa antigen
INVENTOR(S): Cripps, Allan William; Kyd, Jannelle; Dunkley,
Margaret; Clancy, Robert Llewellyn
PATENT ASSIGNEE(S): Auspharm International Limited, Australia;
Chapman, Paul William
SOURCE: PCT Int. Appl., 23 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 2
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9832769	A1	19980730	WO 1998-GB217	19980126
W: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, SD, SZ, UG, ZW, AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG				
ZA 9800587	A	19990723	ZA 1998-587	19980123
AU 9857717	A1	19980818	AU 1998-57717	19980126
EP 980389	A1	20000223	EP 1998-901378	19980126
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, FI				
JP 2001511125	T2	20010807	JP 1998-531741	19980126
PRIORITY APPLN. INFO.: GB 1997-1489 A 19970124 WO 1998-GB217 W 19980126				
AB	A novel antigen from P. aeruginosa is provided. The use of the antigen in detecting/diagnosing P. aeruginosa as well as its use in eliciting an immune response are also provided.			
IT	211046-36-5D, derivs. RL: PRP (Properties); THU (Therapeutic use); BIOL (Biological study); USES (Uses) (Pseudomonas aeruginosa antigen for diagnosing and treating Pseudomonas aeruginosa infection in cystic fibrosis patients)			
RN	211046-36-5 HCAPLUS			
CN	Peptide, (Glu-Glu-Lys-Thr-Pro-Leu-Thr-Thr-Ala-Ala-Xaa-Ala-Pro-Val-Val-Xaa-Asn-Ala) (9CI) (CA INDEX NAME)			

SEQ 1 EEKTPLTAA XAPVVXNA

Searcher : Shears 571-272-2528

1-24-97

THIS PAGE BLANK (USPTO)

09/359426

REFERENCE COUNT: 3 THERE ARE 3 CITED REFERENCES AVAILABLE FOR
THIS RECORD. ALL CITATIONS AVAILABLE IN
THE RE FORMAT

=> fil hom
FILE 'HOME' ENTERED AT 15:45:23 ON 29 JAN 2004

THIS PAGE BLANK (USPTO)

Devis.
09/359426 Page 1
Seq. ID 2 w/
Interf

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 13:09:30 ; Search time 171 Seconds
(without alignments)
101.102 Million cell updates/sec

Title: US-09-359-426C-2

Perfect score: 81

Sequence: 1 XEETPLTAAAXAPVVXNA 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 5728757 seqs, 909918778 residues

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA Main:*

- 1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/1/paa/US097A_COMB.pep.*
- 22: /cgn2_6/ptodata/1/paa/US097B_COMB.pep.*
- 23: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
- 24: /cgn2_6/ptodata/1/paa/US099A_COMB.pep.*
- 25: /cgn2_6/ptodata/1/paa/US099B_COMB.pep.*
- 26: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
- 27: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
- 28: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
- 29: /cgn2_6/ptodata/1/paa/US103_COMB.pep.*
- 30: /cgn2_6/ptodata/1/paa/US104_COMB.pep.*
- 31: /cgn2_6/ptodata/1/paa/US106_COMB.pep.*
- 32: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	76	93.8	19 17	US-09-359-426C-2 Sequence 2, Appli

2	63	77.8	482	22	US-09-791-537-125872	Sequence 125872, A
3	63	77.8	482	29	US-10-366-683-21327	Sequence 21327, A
4	63	77.8	482	30	US-10-419-128-21327	Sequence 21327, A
5	53	65.4	406	19	US-09-543-681A-4268	Sequence 4268, Ap
6	53	65.4	406	31	US-10-603-114-4268	Sequence 4268, Ap
7	53	65.4	484	22	US-09-791-537-30924	Sequence 30924, A
8	53	65.4	495	28	US-10-282-122A-68774	Sequence 68774, A
9	48	59.3	480	28	US-10-282-122A-77773	Sequence 77773, A
10	45	55.6	128	22	US-09-791-537-19306	Sequence 19306, A
11	45	55.6	478	29	US-10-369-493-13806	Sequence 13806, A
12	45	55.6	478	32	US-60-360-039-13806	Sequence 13806, A
13	44	54.3	484	22	US-09-791-537-16868	Sequence 16868, A
14	44	54.3	104	30	US-10-437-963-199308	Sequence 199308, A
15	44	54.3	291	19	US-09-513-996A-35912	Sequence 35912, A
16	44	54.3	291	21	US-09-708-427-23365	Sequence 23365, A
17	44	54.3	292	19	US-09-513-996A-35911	Sequence 35911, A
18	44	54.3	292	21	US-09-708-427-23364	Sequence 23364, A
19	44	54.3	306	26	US-10-049-750-12	Sequence 12, Appl
20	44	54.3	306	28	US-10-238-075-1139	Sequence 1139, Ap
21	44	54.3	427	19	US-09-513-996A-35910	Sequence 35910, A
22	44	54.3	427	21	US-09-708-427-23363	Sequence 23363, A
23	44	54.3	427	29	US-10-361-942-346	Sequence 346, App
24	44	54.3	427	32	US-60-356-051-1468	Sequence 1468, Ap
25	44	54.3	584	28	US-10-282-122A-47689	Sequence 47689, A
26	43	53.1	373	30	US-10-437-963-141172	Sequence 141172, A
27	43	53.1	483	22	US-09-791-537-101743	Sequence 101743, A
28	43	53.1	483	27	US-10-156-761-10589	Sequence 10589, A
29	43	53.1	506	22	US-09-791-537-92565	Sequence 92565, A
30	42	51.9	120	21	US-09-708-427-80045	Sequence 80045, A
31	42	51.9	138	21	US-09-708-427-80044	Sequence 80044, A
32	42	51.9	147	21	US-09-708-427-80043	Sequence 80043, A
33	42	51.9	161	21	US-09-708-427-27615	Sequence 27615, A
34	42	51.9	172	19	US-09-513-996A-54519	Sequence 54519, A
35	42	51.9	204	21	US-09-708-427-27614	Sequence 27614, A
36	42	51.9	215	19	US-09-513-996A-54518	Sequence 54518, A
37	42	51.9	321	21	US-09-708-427-27613	Sequence 27613, A
38	42	51.9	322	28	US-10-219-999-51870	Sequence 51870, A
39	42	51.9	322	30	US-10-425-114-63257	Sequence 63257, A
40	42	51.9	332	19	US-09-513-996A-54517	Sequence 54517, A
41	42	51.9	351	19	US-09-573-655A-1483	Sequence 1483, Ap
42	42	51.9	351	19	US-09-573-655B-1483	Sequence 1483, Ap
43	42	51.9	487	22	US-09-791-537-56402	Sequence 56402, A
44	41	50.6	132	19	US-09-566-394-22	Sequence 22, Appl
45	41	50.6	132	19	US-09-566-394-198	Sequence 198, App

ALIGNMENTS

RESULT 1
US-09-359-426C-2
; Sequence 2, Application US/09359426C
; GENERAL INFORMATION:
; APPLICANT: Cripps, Alan William
; APPLICANT: Clancy, Robert Llewellyn
; APPLICANT: Dunkley, Margaret
; TITLE OF INVENTION: Antigen
; FILE REFERENCE: A33655 064727.0105
; CURRENT APPLICATION NUMBER: US/09/359,426C
; CURRENT FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: PCT/GB98/00217
; PRIOR FILING DATE: 1998-01-26
; PRIOR APPLICATION NUMBER: GB 9701489.8
; PRIOR FILING DATE: 1997-01-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 19
; TYPE: PRT
; ORGANISM: P. Aeruginosa
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1)...(19)

APPLICANT: GARY BRETON


```

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7773
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Versinia pestis
US-10-282-122A-7773

Query Match      59.3%; Score 48; DB 28; Length 480;
Best Local Similarity 68.8%; Pred. No. 26;
Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 3 EKTPLTTAAXAPVXN 18
Db 4 KKKGLTTAAGAPVVDN 19

RESULT 10
US-09-791-537-19306
; Sequence 19306, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19306
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Streptomyces antibioticus
US-09-791-537-19306

Query Match      55.6%; Score 45; DB 22; Length 128;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 EEKTLTTAAXAPV 16
Db 28 EEKFDVTAAGAPV 42

RESULT 11
US-10-369-493-13806
; Sequence 13806, Application US/10369493
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13806
; LENGTH: 478
; TYPE: PRT

```

```

; ORGANISM: Pseudomonas fluorescens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(478)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-13806

Query Match      55.6%; Score 45; DB 29; Length 478;
Best Local Similarity 83.3%; Pred. No. 89;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 7 LTTAAXAPVXN 18
Db 1 LTTAAGAPVVDN 12

RESULT 12
US-60-360-039-13806
; Sequence 13806, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13806
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Pseudomonas fluorescens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(478)
; OTHER INFORMATION: unsure at all Xaa locations
US-60-360-039-13806

Query Match      55.6%; Score 45; DB 32; Length 478;
Best Local Similarity 83.3%; Pred. No. 89;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 7 LTTAAXAPVXN 18
Db 1 LTTAAGAPVVDN 12

RESULT 13
US-09-791-537-16868
; Sequence 16868, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16868
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Actinobacillus actinomycetemcomitans
US-09-791-537-16868

Query Match      55.6%; Score 45; DB 22; Length 484;
Best Local Similarity 83.3%; Pred. No. 90;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 7 LTTAAXAPVXN 18
|||||
Db 14 LTTAAGAPVDN 25

RESULT 14
US-10-437-963-199308
; Sequence 199308, Application US/10437963
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 199308
; LENGTH: 104
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_94886C.1.pap
US-10-437-963-199308

Query Match 54.3%; Score 44; DB 30; Length 104;
Best Local Similarity 69.2%; Pred. No. 20;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKTPLTTAAXAP 14
|||||
Db 79 EKTPLVTPOGAP 91

RESULT 15
US-09-513-996A-35912
; Sequence 35912, Application US/09513996A
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-709P
; CURRENT APPLICATION NUMBER: US/09/513,996A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 35912
; LENGTH: 291
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; NAME/KEY: UNSURE
; LOCATION: 1..291
; OTHER INFORMATION: any n or Xaa = unknown
; FEATURE:
; OTHER INFORMATION: Location 1..291 / Ceres Seq. ID 1808522
US-09-513-996A-35912

Query Match 54.3%; Score 44; DB 19; Length 291;
Best Local Similarity 60.0%; Pred. No. 72;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKTPLTTAAXAPV 16
|||||
Db 72 EKAEMTTAMQSPV 86

RESULT 16
US-09-708-427-23365
; Sequence 23365, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23365
; LENGTH: 291
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; NAME/KEY: misc_feature
; LOCATION: 1..291
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..291
; OTHER INFORMATION: Ceres Seq. ID 1808522
US-09-708-427-23365

Query Match 54.3%; Score 44; DB 21; Length 291;
Best Local Similarity 60.0%; Pred. No. 72;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKTPLTTAAXAPV 16
|||||
Db 72 EKAEMTTAMQSPV 86

RESULT 17
US-09-513-996A-35911
; Sequence 35911, Application US/09513996A
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-709P
; CURRENT APPLICATION NUMBER: US/09/513,996A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 35911
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; NAME/KEY: UNSURE
; LOCATION: 1..292
; OTHER INFORMATION: any n or Xaa = unknown
; FEATURE:
; OTHER INFORMATION: Location 1..292 / Ceres Seq. ID 1808521
US-09-513-996A-35911

Query Match 54.3%; Score 44; DB 19; Length 292;
Best Local Similarity 60.0%; Pred. No. 73;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKTPLTTAAXAPV 16
|||||
Db 73 EKAEMTTAMQSPV 87

RESULT 18
US-09-708-427-23364
; Sequence 23364, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FEATURE:
; OTHER INFORMATION: THEREBY

FILE REFERENCE: 2750-1243P
CURRENT APPLICATION NUMBER: US/09/708,427
CURRENT FILING DATE: 2000-11-09
NUMBER OF SEQ ID NOS: 85364
SOFTWARE: PatentIn version 3.1
SEQ ID NO 23364
LENGTH: 292
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..292
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc.feature
LOCATION: 1..292
OTHER INFORMATION: Ceres Seq. ID 1808521
US-09-708-427-23364

Query Match 54.3%; Score 44; DB 21; Length 292;
Best Local Similarity 60.0%; Pred. No. 73;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAPV 16
||| :||| :|||
Db 73 EEKAEVTTAMQSPV 87

RESULT 19
US-10-049-750-12
Sequence 12, Application US/10049750
GENERAL INFORMATION:
APPLICANT: Fischer, Wilhelm
APPLICANT: Inlenfeldt, Hans-Georg
APPLICANT: Barzu, Octavian
APPLICANT: Sakamoto, Hiroshi
APPLICANT: Pistotnik, Elisabeth
APPLICANT: Marliere, Philippe
APPLICANT: Pochet, Sylvie
TITLE OF INVENTION: Enzymatic synthesis of deoxyribonucleosides
FILE REFERENCE: 2037PMO Deoxyribonucleosides
CURRENT APPLICATION NUMBER: US/10/049,750
CURRENT FILING DATE: 2002-02-15
PRIOR APPLICATION NUMBER: EP99116425.2
PRIOR FILING DATE: 1999-08-20
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 12
LENGTH: 306
TYPE: PRT
ORGANISM: Salmonella typhi
US-10-049-750-12

Query Match 54.3%; Score 44; DB 26; Length 306;
Best Local Similarity 47.1%; Pred. No. 77;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 EKTPLTTAAAPVXNA 19
||| :||| :|||
Db 84 EKVPCTSSGVAFVNA 100

RESULT 20
US-10-238-075-1139
Sequence 1139, Application US/10238075
GENERAL INFORMATION:
APPLICANT: I.N.S.E.R.M.
TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated from E. coli and biological uses of these polynucleotides and of their derivatives
FILE REFERENCE: BLANDINE
CURRENT APPLICATION NUMBER: US/10/238,075
CURRENT FILING DATE: 2002-09-10
PRIOR APPLICATION NUMBER: 0003145
PRIOR FILING DATE: 2000-03-10

NUMBER OF SEQ ID NOS: 1576
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1139
LENGTH: 306
TYPE: PRT
ORGANISM: Escherichia coli
US-10-238-075-1139

Query Match 54.3%; Score 44; DB 28; Length 306;
Best Local Similarity 47.1%; Pred. No. 77;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 EKTPLTTAAAPVXNA 19
||| :||| :|||
Db 84 EKVPCTSSGVAFVNA 100

RESULT 21
US-09-513-996A-35910
Sequence 35910, Application US/09513996A
GENERAL INFORMATION:
APPLICANT: N. ALEXANDROV et al.
TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
FILE REFERENCE: 2750-709P
CURRENT APPLICATION NUMBER: US/09/513,996A
CURRENT FILING DATE: 2000-02-25
NUMBER OF SEQ ID NOS: 81028
SEQ ID NO 35910
LENGTH: 427
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: UNSURE
LOCATION: 1..427
OTHER INFORMATION: any n or Xaa = unknown
FEATURE:
OTHER INFORMATION: Location 1..427 / Ceres Seq. ID 1808520
US-09-513-996A-35910

Query Match 54.3%; Score 44; DB 19; Length 427;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAPV 16
||| :||| :|||
Db 208 EEKAEVTTAMQSPV 222

RESULT 22
US-09-708-427-23363
Sequence 23363, Application US/09708427
GENERAL INFORMATION:
APPLICANT: N. ALEXANDROV et al.
TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
FILE REFERENCE: 2750-1243P
CURRENT APPLICATION NUMBER: US/09/708,427
CURRENT FILING DATE: 2000-11-09
NUMBER OF SEQ ID NOS: 85364
SOFTWARE: PatentIn version 3.1
SEQ ID NO 23363
LENGTH: 427
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..427
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc.feature
LOCATION: 1..427
OTHER INFORMATION: Ceres Seq. ID 1808520
US-09-708-427-23363


```
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT MRT4530_42301C.1.pep
US-10-437-963-141172

Query Match      53.1%; Score 43; DB 30; Length 373;
Best Local Similarity 61.5%; Pred. No. 1.5e+02;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAP 14
Db 222 DEETPTTTAAAP 234

RESULT 27
US-09-791-537-101743
; Sequence 101743, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 101743
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Streptomyces violaceus
US-09-791-537-101743

Query Match      53.1%; Score 43; DB 22; Length 483;
Best Local Similarity 69.2%; Pred. No. 2e+02;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 6 PLTTAAXAPVVXN 18
Db 5 PLTTEAGAPVDN 17

RESULT 28
US-10-156-761-10589
; Sequence 10589, Application US/10156761
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10589
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10589

Query Match      53.1%; Score 43; DB 27; Length 483;
Best Local Similarity 69.2%; Pred. No. 2e+02;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 6 PLTTAAXAPVVXN 18
Db 5 PLTTEAGAPVDN 17

RESULT 29
US-09-791-537-92565
; Sequence 92565, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 92565
; LENGTH: 506
; TYPE: PRT
; ORGANISM: Pseudomonas fluorescens
US-09-791-537-92565

Query Match      53.1%; Score 43; DB 22; Length 506;
Best Local Similarity 71.4%; Pred. No. 2.2e+02;
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 TPLTTAAXAPVVXN 18
Db 11 TLTTEAGAPVDN 24

RESULT 30
US-09-708-427-80045
; Sequence 80045, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 80045
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..120
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc.feature
; LOCATION: 1..120
; OTHER INFORMATION: Ceres Seq. ID 1957331
US-09-708-427-80045

Query Match      51.9%; Score 42; DB 21; Length 120;
Best Local Similarity 50.0%; Pred. No. 55;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EKTPLTTAAXAPVVXNA 19
Db 79 EKTPEKKTGVPVVKKA 96

RESULT 31
US-09-708-427-80044
; Sequence 80044, Application US/09708427
; GENERAL INFORMATION:
```

; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 80044
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; NAME/KEY: misc.feature
; LOCATION: 1..138
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc.feature
; LOCATION: 1..138
; OTHER INFORMATION: Ceres Seq. ID 1957330
US-09-708-427-80044

Query Match 51.9%; Score 42; DB 21; Length 138;
Best Local Similarity 50.0%; Pred. No. 65;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAPVVXNA 19
|||||
Db 97 EKTPEKKTGVPVVKKA 114

RESULT 32
US-09-708-427-80043
; Sequence 80043, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 80043
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; NAME/KEY: misc.feature
; LOCATION: 1..147
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc.feature
; LOCATION: 1..147
; OTHER INFORMATION: Ceres Seq. ID 1957329
US-09-708-427-80043

Query Match 51.9%; Score 42; DB 21; Length 147;
Best Local Similarity 50.0%; Pred. No. 71;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAPVVXNA 19
|||||
Db 106 EKTPEKKTGVPVVKKA 123

RESULT 33
US-09-708-427-27615
; Sequence 27615, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427

; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27615
; LENGTH: 161
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; NAME/KEY: misc.feature
; LOCATION: 1..161
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc.feature
; LOCATION: 1..161
; OTHER INFORMATION: Ceres Seq. ID 1821333
US-09-708-427-27615

Query Match 51.9%; Score 42; DB 21; Length 161;
Best Local Similarity 50.0%; Pred. No. 79;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAPVVXNA 19
|||||
Db 131 EKTPEKKTGVPVVKKA 148

RESULT 34
US-09-513-996A-54519
; Sequence 54519, Application US/09513996A
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-709P
; CURRENT APPLICATION NUMBER: US/09/513,996A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 54519
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; NAME/KEY: UNSURE
; LOCATION: 1..172
; OTHER INFORMATION: any n or Xaa = unknown
; FEATURE:
; OTHER INFORMATION: Location 1..172 / Ceres Seq. ID 2109050
US-09-513-996A-54519

Query Match 51.9%; Score 42; DB 19; Length 172;
Best Local Similarity 50.0%; Pred. No. 86;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAPVVXNA 19
|||||
Db 131 EKTPEKKTGVPVVKKA 148

RESULT 35
US-09-708-427-27614
; Sequence 27614, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27614
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana

```

;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..204
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc feature
; LOCATION: 1..204
; OTHER INFORMATION: Ceres Seq. ID 1821332
US-09-708-427-27614

Query Match          51.9%; Score 42; DB 21; Length 204;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EKETPLTTAAAXAPVVXNA 19
    |||||: |||
Db 174 EKETPVEKKTGVPVVKKA 191

RESULT 36
US-09-513-996A-54518
; Sequence 54518, Application US/09513996A
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-709P
; CURRENT APPLICATION NUMBER: US/09/513,996A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 54518
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 1..215
; OTHER INFORMATION: any n or Xaa = unknown
; FEATURE:
; OTHER INFORMATION: Location 1..215 / Ceres Seq. ID 2109049
US-09-513-996A-54518

Query Match          51.9%; Score 42; DB 19; Length 215;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EKETPLTTAAAXAPVVXNA 19
    |||||: |||
Db 174 EKETPVEKKTGVPVVKKA 191

RESULT 37
US-09-708-427-27613
; Sequence 27613, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27613
; LENGTH: 321
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..321
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc feature
; LOCATION: 1..321
; OTHER INFORMATION: Ceres Seq. ID 1821331

```

```

US-09-708-427-27613

Query Match          51.9%; Score 42; DB 21; Length 321;
Best Local Similarity 50.0%; Pred. No. 1.9e+02;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EKETPLTTAAAXAPVVXNA 19
    |||||: |||
Db 291 EKETPVEKKTGVPVVKKA 308

RESULT 38
US-10-219-999-51870
; Sequence 51870, Application US/10219999
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Stein, Joshua
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-10(52726)C
; CURRENT APPLICATION NUMBER: US/10/219,999
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/324,109
; PRIOR FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/312,544
; PRIOR FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 63520
; SEQ ID NO 51870
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Zea mays
US-10-219-999-51870

Query Match          51.9%; Score 42; DB 28; Length 322;
Best Local Similarity 50.0%; Pred. No. 1.9e+02;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EKETPLTTAAAXAPVVXNA 19
    |||||: |||
Db 44 EETPTAATAAASPAPDRA 61

RESULT 39
US-10-425-114-63257
; Sequence 63257, Application US/10425114
; GENERAL INFORMATION:
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 63257
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3180-051-B8_FLI.pep
US-10-425-114-63257

Query Match          51.9%; Score 42; DB 30; Length 322;
Best Local Similarity 50.0%; Pred. No. 1.9e+02;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

```

QY 2 EKTPLTTAAAPVXNA 19
|||:|
Db 44 EETPAATAASPAPDRA 61

RESULT 40
US-09-513-996A-54517
; Sequence 54517, Application US/09513996A
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THEREBY
; FILE REFERENCE: 2750-709P
; CURRENT APPLICATION NUMBER: US/09/513,996A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 54517
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 1..332
; OTHER INFORMATION: any n or Xaa = unknown
; FEATURE:
; OTHER INFORMATION: Location 1..332 / Ceres Seq. ID 2109048
US-09-513-996A-54517

Query Match 51.9%; Score 42; DB 19; Length 332;
Best Local Similarity 50.0%; Pred. No. 1.9e+02;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

-QY 2 EKTPLTTAAAPVXNA 19
|||:|
Db 291 EKTPEKKTGVVVKKA 308

-Search completed: January 28, 2004, 13:14:54
Job time : 172 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 13:08:39 ; Search time 22 Seconds
(without alignments)
64.984 Million cell updates/sec

Title: US-09-359-426C-2

Perfect score: 81
Sequence: 1 XEKTPLTAAAPVXNA 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 364858 seqs, 75244621 residues

Total number of hits satisfying chosen parameters: 364858

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA.New.*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	51.9	322	5	US-10-425-114A-63257, A
2	41	50.6	1012	5	US-09-614-150A-27570, A
3	40	49.4	788	5	US-09-614-150A-27570, A
4	40	49.4	1473	5	US-09-614-150A-30864, A
5	39.5	48.8	1793	5	US-09-614-150A-5631, A
6	39	48.1	484	1	PCT-US03-10308A-29
7	39	48.1	484	6	US-10-406-686A-29
8	39	48.1	486	6	US-10-679-063-21767, A
9	39	48.1	509	6	US-10-679-063-21767, A
10	39	48.1	560	7	US-60-470-920-16
11	38.5	47.5	842	5	US-09-614-150A-30690, A
12	38	46.9	116	6	US-10-425-114A-68517, A
13	38	46.9	186	6	US-10-425-114A-46588, A
14	38	46.9	225	5	US-09-614-150A-41325, A
15	38	46.9	229	6	US-10-425-114A-64721, A
16	38	46.9	1162	7	US-60-490-890-611
17	38	46.9	19695	6	US-10-084-846A-3
18	37	45.7	164	6	US-10-425-114A-59823, A
19	37	45.7	229	6	US-10-425-114A-39997, A
20	37	45.7	304	6	US-10-679-063-19753, A
21	37	45.7	324	6	US-10-739-930-7671, A
22	37	45.7	396	6	US-10-415-182A-6506, A
23	37	45.7	498	5	US-09-323-998E-57
24	37	45.7	500	5	US-09-323-998E-58
25	37	45.7	500	5	US-09-323-998E-59
26	37	45.7	564	6	US-10-679-063-23185, A

ALIGNMENTS

RESULT 1

US-10-425-114A-63257
; Sequence 63257, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 63257
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3180-051-E8_FLI.pep
US-10-425-114A-63257

Query Match 51.9%; Score 42; DB 6; Length 322;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EKTPLTAAAPVXNA 19
||| ||| :|
DB 44 EETPAATAAASPAPDRA 61

RESULT 2

US-09-614-150A-27570
; Sequence 27570, Application US/09614150A
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CLO00728
; CURRENT APPLICATION NUMBER: US/09/614,150A
; CURRENT FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19

27 37 45.7 630 5 US-09-614-150A-17034 Sequence 17034, A
28 37 45.7 715 5 US-10-425-114A-65245 Sequence 65245, A
29 37 45.7 865 5 US-09-614-150A-18177 Sequence 18177, A
30 37 45.7 1064 1 PCT-US03-38193-3242 Sequence 3242, Ap
31 37 45.7 1064 6 US-10-723-860-3242 Sequence 3242, Ap
32 37 45.7 1064 6 US-10-700-439-142 Sequence 142, Ap
33 37 45.7 1064 7 US-60-519-570-229 Sequence 229, Ap
34 37 45.7 3390 6 US-10-719-547-22 Sequence 22, Appl
35 36.5 45.1 314 5 US-09-614-150A-39876 Sequence 39876, A
36 36 44.4 208 5 US-09-614-150A-864 Sequence 864, App
37 36 44.4 242 6 US-10-679-063-14309 Sequence 14309, A
38 36 44.4 262 6 US-10-739-930-9368 Sequence 9368, Ap
39 36 44.4 310 6 US-10-425-114A-52153 Sequence 52153, A
40 36 44.4 318 6 US-10-425-114A-51755 Sequence 51755, A
41 36 44.4 323 6 US-10-425-114A-70539 Sequence 70539, A
42 36 44.4 340 6 US-10-425-114A-66503 Sequence 66503, A
43 36 44.4 486 6 US-10-679-063-19265 Sequence 19265, A
44 36 44.4 490 6 US-10-425-114A-43398 Sequence 43398, A
45 36 44.4 501 5 US-09-323-998E-55 Sequence 55, Appl

```

; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27570
; LENGTH: 1012
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-09-614-150A-27570

```

```

Query Match          50.6%; Score 41; DB 5; Length 1012;
Best Local Similarity 53.3%; Pred. No. 82;
Matches      8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      2 EKTPLTTAAXAPV 16
      ||||| ||||| |||||
Db      695 EVTPYNTAPHSFV 709

```

```

RESULT 3
US-09-614-150A-8541
; Sequence 8541, Application US/09614150A
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/09/614,150A
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8541
; LENGTH: 788
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-09-614-150A-8541

```

```

Query Match          49.4%; Score 40; DB 5; Length 788;
Best Local Similarity 52.9%; Pred. No. 92;
Matches      9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

```

```

QY      3 EKTPLTTAAXAPVXNA 19
      ||||| ||||| |||||
Db      602 EDGPTTTAAAPLASAA 618

```

```

RESULT 4
US-09-614-150A-30864
; Sequence 30864, Application US/09614150A
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/09/614,150A
; CURRENT FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30864
; LENGTH: 1473
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-09-614-150A-30864

```

```

Query Match          49.4%; Score 40; DB 5; Length 1473;
Best Local Similarity 53.3%; Pred. No. 1.8e+02;
Matches      8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      2 EKTPLTTAAXAPV 16
      ||||| ||||| |||||
Db      1315 DEQAPYTEALGPV 1329

```

```

RESULT 5
US-09-614-150A-5631
; Sequence 5631, Application US/09614150A
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/09/614,150A
; CURRENT FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637

```


; PRIOR FILING DATE: 2000-03-23
 ; NUMBER OF SEQ ID NOS: 43008
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5631
 ; LENGTH: 1793
 ; TYPE: PRT
 ; ORGANISM: DROSOPHILA
 US-09-614-150A-5631

Query Match 48.8%; Score 39.5; DB 5; Length 1793;
 Best Local Similarity 52.6%; Pred. No. 2.7e+02;
 Matches 10; Conservative 3; Mismatches 5; Indels 1; Gaps 1;
 QY 2 EKTPLTT-AAAXAPVVXNA 19
 Db 710 KKSSPTTTPAARAPVAQNA 728

RESULT 6
 PCT-US03-10308A-29
 ; Sequence 29, Application PC/TUS0310308A
 ; GENERAL INFORMATION:
 ; APPLICANT: MERIAL LIMITED
 ; TITLE OF INVENTION: ATTENUATED GRAM NEGATIVE BACTERIA
 ; FILE REFERENCE: 454313-3171.1WO
 ; CURRENT APPLICATION NUMBER: PCT/US03/10308A
 ; CURRENT FILING DATE: 2003-08-19
 ; PRIOR APPLICATION NUMBER: 60/370,282
 ; PRIOR FILING DATE: 2002-04-05
 ; NUMBER OF SEQ ID NOS: 118
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 29
 ; TYPE: PRT
 ; LENGTH: 484
 ; ORGANISM: Pasteurella multocida
 PCT-US03-10308A-29

Query Match 48.1%; Score 39; DB 1; Length 484;
 Best Local Similarity 66.7%; Pred. No. 81;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 7 LTAAAXAPVVXN 18
 Db 13 LTNAAGAPIVEN 24

RESULT 7
 US-10-406-686A-29
 ; Sequence 29, Application US/10406686A
 ; GENERAL INFORMATION:
 ; APPLICANT: CROOKE, HELEN RACHEL
 ; APPLICANT: SHEA, JACQUELINE ELIZABETH
 ; APPLICANT: FELDMAN, ROBERT GRAHAM
 ; APPLICANT: COUTEBROZE, SYLVAIN GABRIEL
 ; APPLICANT: LEGROS, FRANCOIS-XAVIER
 ; TITLE OF INVENTION: ATTENUATED GRAM NEGATIVE BACTERIA
 ; FILE REFERENCE: 454313-3171.1
 ; CURRENT APPLICATION NUMBER: US/10/406,686A
 ; CURRENT FILING DATE: 2003-04-03
 ; PRIOR APPLICATION NUMBER: 60/370,282
 ; PRIOR FILING DATE: 2002-04-05
 ; NUMBER OF SEQ ID NOS: 118
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 29
 ; TYPE: PRT
 ; LENGTH: 484
 ; ORGANISM: Pasteurella multocida
 US-10-406-686A-29

Query Match 48.1%; Score 39; DB 6; Length 484;
 Best Local Similarity 66.7%; Pred. No. 81;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 7 LTAAAXAPVVXN 18
 Db 13 LTNAAGAPIVEN 24
 RESULT 8
 US-10-679-063-7762
 ; Sequence 7762, Application US/10679063
 ; GENERAL INFORMATION:
 ; APPLICANT: Edgerton, Michael D
 ; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
 ; FILE REFERENCE: 38-15(52054)B
 ; CURRENT APPLICATION NUMBER: US/10/679,063
 ; CURRENT FILING DATE: 2003-10-02
 ; PRIOR APPLICATION NUMBER: 60/415,758
 ; PRIOR FILING DATE: 2002-10-02
 ; NUMBER OF SEQ ID NOS: 27373
 ; SEQ ID NO 7762
 ; LENGTH: 486
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas sp. CF600
 US-10-679-063-7762

Query Match 48.1%; Score 39; DB 6; Length 486;
 Best Local Similarity 52.9%; Pred. No. 82;
 Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
 QY 3 EKTPLTTAAAXAPVVXNA 19
 Db 181 EETPLTTALLGEVMOAA 197

RESULT 9
 US-10-679-063-21767
 ; Sequence 21767, Application US/10679063
 ; GENERAL INFORMATION:
 ; APPLICANT: Edgerton, Michael D
 ; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
 ; FILE REFERENCE: 38-15(52054)B
 ; CURRENT APPLICATION NUMBER: US/10/679,063
 ; CURRENT FILING DATE: 2003-10-02
 ; PRIOR APPLICATION NUMBER: 60/415,758
 ; PRIOR FILING DATE: 2002-10-02
 ; NUMBER OF SEQ ID NOS: 27373
 ; SEQ ID NO 21767
 ; LENGTH: 509
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis MA-4680
 US-10-679-063-21767

Query Match 48.1%; Score 39; DB 6; Length 509;
 Best Local Similarity 52.9%; Pred. No. 86;
 Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
 QY 3 EKTPLTTAAAXAPVVXNA 19
 Db 189 EITPLTTAALLVLLVEA 205

RESULT 10
 US-60-470-920-16
 ; Sequence 16, Application US/60470920
 ; GENERAL INFORMATION:
 ; APPLICANT: Luke, Catherine, J.
 ; APPLICANT: Vilalta, Adrian
 ; APPLICANT: Wloch, Mary K.
 ; APPLICANT: Evans, Thomas G.
 ; APPLICANT: Geall, Andrew J.
 ; TITLE OF INVENTION: DNA Influenza Vaccine Compositions
 ; FILE REFERENCE: 1530.0640000
 ; CURRENT APPLICATION NUMBER: US/60/470,920
 ; CURRENT FILING DATE: 2003-05-16
 ; NUMBER OF SEQ ID NOS: 45

```

; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Influenza B virus
US-60-470-920-16

```

Query Match 48.1%; Score 39; DB 7; Length 560;
Best Local Similarity 44.4%; Pred. No. 95;
Matches 8; Conservative 3; Mismatches 7; Indels

Oy 2 E E K T P L T T A A X A P V V X N A 19
 || | | | : | | : |
Db 19 E E L T P G T S G A T R P I I K P A 36

```

RESULT 11
US-09-614-150A-30690
; Sequence 30690, Application US/09614150A
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/09/614,150A
; CURRENT FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30690
; LENGTH: 842
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-09-614-150A-30690

```

Query Match	47.5%	Score 38.5;	DB 5;	Length 842;
Best Local Similarity	50.0%;	Pred. No. 1.8e+00;		
Matches	9;	Conservative 3;	Mismatches 5;	Indels 1; Gaps 1;
Qy	3	EKTEL-TTAAAPVVXNA	19	
		:		:
Db	598	EKTRIVSTTKAPVWSTS	615	

RESULT 12
US-10-425-114A-68517
; Sequence 68517, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

```

; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 68517
; LENGTH: 116
; TYPE: PRT
; ORGANISM: Zea mays subsp. mexicana
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZM2RTEOSII
US-10-425-114A-68517

```

Query Match 46.9%; Score 38; DB.6; Length 116;
Best Local Similarity 63.6%; Pred. No.26;
Matches 7: Conservative 3; Mismatches 1; Indels

QY 5 TPLTTAAXAPV 15
:|||||: :||
db 104 SPLTTASLSPV 114

```

RESULT 13
US-10-425-114A-46588
; Sequence 46588, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 46588
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700427386_FLI pep
US-10-425-114A-46588

```

Query Match	46.9%;	Score 38;	DB 6;	Length 186;
Best Local Similarity	63.6%;	Pred. No. 44;		
Matches	7:	Conservative	3:	Mismatches
			1:	Indels
			0:	Gaps
			0:	Gaps

Qy 5 TPLTTAAXAPV 15
:|||||: :||
pb 111 SPLTTASLSBP 121

RESULT 14
US-09-614-150A-41325
; Sequence 41325, Application US/09614150A
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/09/614,150A
; CURRENT FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769

; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41325
; LENGTH: 225
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-09-614-150A-41325

Query Match 46.9%; Score 38; DB 5; Length 225;
Best Local Similarity 58.3%; Pred. No. 53;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 TPLTTAAAXAPVV 16
|||:|||:
Db 201 TPVTPAPTAPVI 212

RESULT 15
US-10-425-114A-64721
; Sequence 64721, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 64721
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LB4605-005-B1_FLI.pep
US-10-425-114A-64721

Query Match 46.9%; Score 38; DB 6; Length 229;
Best Local Similarity 47.1%; Pred. No. 54;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAPVVXN 18
|||:|||:
Db 1 EETPPLLAAGSPVTAS 17

RESULT 16
US-60-490-890-611
; Sequence 611, Application US/60490890
; GENERAL INFORMATION:
; APPLICANT: Li, Martha
; APPLICANT: Rupnow, Brent A.
; APPLICANT: Webster, Kevin R.
; APPLICANT: Jackson, Donald
; APPLICANT: Wong, Tai W.
; TITLE OF INVENTION: BIOMARKERS OF CYCLIN-DEPENDENT KINASE MODULATION
; FILE REFERENCE: D0310 PSP
; CURRENT APPLICATION NUMBER: US/60/490,890
; CURRENT FILING DATE: 2003-07-29
; NUMBER OF SEQ ID NOS: 2779

; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 611
; LENGTH: 1162
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-490-890-611

Query Match 46.9%; Score 38; DB 7; Length 1162;
Best Local Similarity 46.7%; Pred. No. 3e+02;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 TPLTTAAAXAPVVXNA 19
|||:|||:
Db 550 TPRPAAASPVSVA 564

RESULT 17
US-10-084-846A-3
; Sequence 3, Application US/10084846A
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUELENWEG, AGNES
; APPLICANT: TREPZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 3
; LENGTH: 19695
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 1: amino acid sequence encoded by coding strand 1.
; OTHER INFORMATION: Start codon: 99a, Start position: nucleotide 1.
US-10-084-846A-3

Query Match 46.9%; Score 38; DB 6; Length 19695;
Best Local Similarity 58.3%; Pred. No. 6e+03;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 TPLTTAAAXAPVV 16
|||:|||:
Db 15780 TPASTRAAAPVI 15791

RESULT 18
US-10-425-114A-59823
; Sequence 59823, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 59823
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:

; OTHER INFORMATION: Clone ID: LIB3912-049-F8_FLI.pep
US-10-425-114A-59823

Query Match 45.7%; Score 37; DB 6; Length 164;
Best Local Similarity 80.0%; Pred. No. 56;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 TPLTTAAAXAP 14
||| |||||
Db 30 TPSTTAAPAP 39

RESULT 19

US-10-425-114A-39997

; Sequence 39997, Application US/10425114A

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(5313)B

; CURRENT APPLICATION NUMBER: US/10/425,114A

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 39997

; LENGTH: 229

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: 700945759_FLI.pep

US-10-425-114A-39997

Query Match 45.7%; Score 37; DB 6; Length 229;

Best Local Similarity 46.2%; Pred. No. 80;

Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 BEKTPLTAAAXAP 14

||| |||||

Db 194 KEKGPMTTASP 206

RESULT 20

US-10-679-063-19753

; Sequence 19753, Application US/10679063

; GENERAL INFORMATION:

; APPLICANT: Edgerton, Michael D

; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES

; FILE REFERENCE: 38-15(52054)B

; CURRENT APPLICATION NUMBER: US/10/679,063

; CURRENT FILING DATE: 2003-10-02

; PRIOR APPLICATION NUMBER: 60/415,758

; PRIOR FILING DATE: 2002-10-02

; NUMBER OF SEQ ID NOS: 27373

; SEQ ID NO 19753

; LENGTH: 304

; TYPE: PRT

; ORGANISM: Bradyrhizobium japonicum USDA 110

US-10-679-063-19753

Query Match 45.7%; Score 37; DB 6; Length 304;

Best Local Similarity 50.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 TPLTTAAAXAPVXN 18

||| |||||

Db 106 TPLWVAIAAPLIGN 119

RESULT 21

US-10-739-930-7671

; Sequence 7671, Application US/10739930

; GENERAL INFORMATION:

; APPLICANT: Kovalic, David K.

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH

; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT

; FILE REFERENCE: 38-21(53377)B

; CURRENT APPLICATION NUMBER: US/10/739,930

; CURRENT FILING DATE: 2003-12-18

; NUMBER OF SEQ ID NOS: 11088

; SEQ ID NO 7671

; LENGTH: 324

; TYPE: PRT

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: ZENMA-23APR03-C2255_1.P

US-10-739-930-7671

Query Match 45.7%; Score 37; DB 6; Length 324;

Best Local Similarity 53.8%; Pred. No. 1.2e+02;

Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 BEKTPLTAAAXAP 14

: ||| |||||

Db 37 ORKAPATTAALP 49

RESULT 22

US-10-415-182A-6506

; Sequence 6506, Application US/10415182A

; GENERAL INFORMATION:

; APPLICANT: Telford, John

; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM STREPTOCOCCUS GROUPS A & B

; FILE REFERENCE:

; CURRENT APPLICATION NUMBER: US/10/415,182A

; CURRENT FILING DATE: 2003-12-09

; PRIOR APPLICATION NUMBER: GB-0026333.5

; PRIOR FILING DATE: 2000-10-27

; PRIOR APPLICATION NUMBER: GB-0028727.6

; PRIOR FILING DATE: 2000-11-24

; PRIOR APPLICATION NUMBER: GB-0105640.7

; PRIOR FILING DATE: 2001-03-07

; NUMBER OF SEQ ID NOS: 12024

; SOFTWARE: SeqWin99, version 1.02

; SEQ ID NO 6506

; LENGTH: 396

; TYPE: PRT

; ORGANISM: Streptococcus agalactiae

US-10-415-182A-6506

Query Match 45.7%; Score 37; DB 6; Length 396;

Best Local Similarity 40.0%; Pred. No. 1.4e+02;

Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 BEKTPLTAAAXAPVV 16

||| |||||

Db 82 EQKYPVLSALSRPII 96

RESULT 23

US-09-323-998E-57

; Sequence 57, Application US/09323998E

; GENERAL INFORMATION:

; APPLICANT: CUNNINGHAM JR., FRANCIS X.

; APPLICANT: SUN, ZAIEN

; TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND

; TITLE OF INVENTION: METHODS OF USE THEREOF

; FILE REFERENCE: 108172-09019

; CURRENT APPLICATION NUMBER: US/09/323,998E

; CURRENT FILING DATE: 1999-06-02

; PRIOR APPLICATION NUMBER: 09/088,724

; PRIOR FILING DATE: 1998-06-02

; PRIOR APPLICATION NUMBER: 09/088,725

; PRIOR FILING DATE: 1998-06-02
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 57
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Capsicum annuum
US-09-323-998E-57

Query Match 45.7%; Score 37; DB 5; Length 498;
Best Local Similarity 72.7%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 9 TAAAXAPVVXNA 19
| | | | | | |
DB 370 TLAAPVVANA 380

RESULT 24
US-09-323-998E-58
; Sequence 58, Application US/09323998E
; GENERAL INFORMATION:
; APPLICANT: CUNNINGHAM JR., FRANCIS X.
; APPLICANT: SUN, ZAIREN
; TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
; FILE REFERENCE: 108172-09019
; CURRENT APPLICATION NUMBER: US/09/323,998E
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: 09/088,724
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 09/088,725
; PRIOR FILING DATE: 1998-06-02
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 58
; TYPE: PRT
; ORGANISM: Lycopersicon esculentum
US-09-323-998E-58

Query Match 45.7%; Score 37; DB 5; Length 500;
Best Local Similarity 72.7%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 9 TAAAXAPVVXNA 19
| | | | | | |
DB 372 TLAAPVVANA 382

RESULT 25
US-09-323-998E-59
; Sequence 59, Application US/09323998E
; GENERAL INFORMATION:
; APPLICANT: CUNNINGHAM JR., FRANCIS X.
; APPLICANT: SUN, ZAIREN
; TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
; FILE REFERENCE: 108172-09019
; CURRENT APPLICATION NUMBER: US/09/323,998E
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: 09/088,724
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 09/088,725
; PRIOR FILING DATE: 1998-06-02
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 59
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Nicotiana tabacum
US-09-323-998E-59

Query Match 45.7%; Score 37; DB 5; Length 500;
Best Local Similarity 72.7%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 9 TAAAXAPVVXNA 19
| | | | | | |
DB 372 TLAAPVVANA 382

RESULT 26
US-10-679-063-23185
; Sequence 23185, Application US/10679063
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52054)B
; CURRENT APPLICATION NUMBER: US/10/679,063
; CURRENT FILING DATE: 2003-10-02
; PRIOR APPLICATION NUMBER: 60/415,758
; PRIOR FILING DATE: 2002-10-02
; NUMBER OF SEQ ID NOS: 27373
; SEQ ID NO 23185
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Oryza sativa (japonica cultivar-group)
US-10-679-063-23185

Query Match 45.7%; Score 37; DB 6; Length 564;
Best Local Similarity 66.7%; Pred. No. 2.1e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 8 TTAAXAPVVXNA 19
| | | | | | |
DB 59 TTSAAPVVVEA 70

RESULT 27
US-09-614-150A-17034
; Sequence 17034, Application US/09614150A
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/09/614,150A
; CURRENT FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17034
; LENGTH: 630
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-09-614-150A-17034

Query Match 45.7%; Score 37; DB 5; Length 630;

```

; TYPE: PRT
; ORGANISM: DROSOPHILA
US-09-614-150A-18177

Query Match 45.7%; Score 37; DB 5; Length 865;
Best Local Similarity 53.3%; Pred. No. 3.3e+02;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAPV 16
Db 102 KDATPATTAAAPLPV 116

RESULT 30
PCT-US03-38193-3242
; Sequence 3242, Application PC/TUS0338193
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.00PC00
; CURRENT APPLICATION NUMBER: PCT/US03/38193
; CURRENT FILING DATE: 2003-11-26
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 3242
; LENGTH: 1064
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US03-38193-3242

Query Match 45.7%; Score 37; DB 1; Length 1064;
Best Local Similarity 61.5%; Pred. No. 4.1e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAP 14
Db 1048 ESSTGTAAASMP 1060

RESULT 31
US-10-723-860-3242
; Sequence 3242, Application US/10723860
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 3242
; LENGTH: 1064
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-3242

Query Match 45.7%; Score 37; DB 6; Length 1064;
Best Local Similarity 61.5%; Pred. No. 4.1e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAP 14
Db 1048 ESSTGTAAASMP 1060

```

```

; TYPE: PRT
; ORGANISM: DROSOPHILA
US-09-614-150A-18177

Query Match 45.7%; Score 37; DB 5; Length 865;
Best Local Similarity 53.3%; Pred. No. 3.3e+02;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAPV 16
Db 102 KDATPATTAAAPLPV 116

RESULT 30
PCT-US03-38193-3242
; Sequence 3242, Application PC/TUS0338193
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.00PC00
; CURRENT APPLICATION NUMBER: PCT/US03/38193
; CURRENT FILING DATE: 2003-11-26
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 3242
; LENGTH: 1064
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US03-38193-3242

Query Match 45.7%; Score 37; DB 1; Length 1064;
Best Local Similarity 61.5%; Pred. No. 4.1e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAP 14
Db 1048 ESSTGTAAASMP 1060

RESULT 31
US-10-723-860-3242
; Sequence 3242, Application US/10723860
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 3242
; LENGTH: 1064
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-3242

Query Match 45.7%; Score 37; DB 6; Length 1064;
Best Local Similarity 61.5%; Pred. No. 4.1e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAP 14
Db 1048 ESSTGTAAASMP 1060

```

```
RESULT 32
US-10-700-439-142
; Sequence 142, Application US/10700439
; GENERAL INFORMATION:
; APPLICANT: Mayo Foundation for Medical Education and Research
; APPLICANT: Bayer Healthcare LLC
; TITLE OF INVENTION: ATTENUATING DENGUE VIRUSES AND CHIMERIC DENGUE VIRUSES
; FILE REFERENCE: NIH214.001C1
; APPLICANT: Burgess, Christopher
; APPLICANT: Myerow, Susan
; APPLICANT: Thiagalingam, Arunthathi
; APPLICANT: Maimonis, Peter
; APPLICANT: Molino, Gary
; APPLICANT: Burgart, Lawrence
; APPLICANT: Boardman, Lisa A.
; APPLICANT: Thibodeau, Steven
; APPLICANT: Lewis, Marcia
; TITLE OF INVENTION: Use of Differentially Expressed Nucleic Acid Sequences as
; TITLE OF INVENTION: Biomarkers for Cancer
; FILE REFERENCE: 1657/2022
; CURRENT APPLICATION NUMBER: US/10/700,439
; CURRENT FILING DATE: 2003-11-04
; NUMBER OF SEQ ID NOS: 186
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 142
; LENGTH: 1064
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-700-439-142

Query Match 45.7%; Score 37; DB 6; Length 1064;
Best Local Similarity 61.5%; Pred. No. 4.1e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 EKTPLTTAAXAP 14
| | | | |
Db 1048 ESSTPGTTAASMP 1060

RESULT 33
US-60-519-270-229
; Sequence 229, Application US/60519270
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: BEGOVICH, Ann
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001490
; CURRENT APPLICATION NUMBER: US/60/519,270
; CURRENT FILING DATE: 2003-11-13
; NUMBER OF SEQ ID NOS: 14048
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 229
; LENGTH: 1064
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-519-270-229

Query Match 45.7%; Score 37; DB 7; Length 1064;
Best Local Similarity 61.5%; Pred. No. 4.1e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 EKTPLTTAAXAP 14
| | | | |
Db 1048 ESSTPGTTAASMP 1060

RESULT 34
US-10-719-547-22
; Sequence 22, Application US/10719547
; GENERAL INFORMATION:
; APPLICANT: Whitehead, Stephen S.
; APPLICANT: Murphy, Brian R.
```

```
; APPLICANT: Hanley, Kathryn A.
; APPLICANT: Blaney, Joseph E. Jr.
; TITLE OF INVENTION: DEVELOPMENT OF MUTATIONS USEFUL FOR
; TITLE OF INVENTION: ATTENUATING DENGUE VIRUSES AND CHIMERIC DENGUE VIRUSES
; FILE REFERENCE: NIH214.001C1
; CURRENT APPLICATION NUMBER: US/10/719,547
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US02/16308
; PRIOR FILING DATE: 2002-05-22
; PRIOR APPLICATION NUMBER: US 60/293049
; PRIOR FILING DATE: 2001-05-22
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 3390
; TYPE: PRT
; ORGANISM: Dengue 3 virus strain H87
US-10-719-547-22

Query Match 45.7%; Score 37; DB 6; Length 3390;
Best Local Similarity 53.8%; Pred. No. 1.4e+03;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKTPLTTAAXAP 14
| | | | |
Db 3315 EDKTPVTWEDVP 3327
```

```
RESULT 35
US-09-614-150A-39876
; Sequence 39876, Application US/09614150A
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/09/614,150A
; CURRENT FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39876
; LENGTH: 314
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-09-614-150A-39876
```

```
Query Match 45.1%; Score 36.5; DB 5; Length 314;
Best Local Similarity 47.8%; Pred. No. 1.4e+02;
Matches 11; Conservative 0; Mismatches 7; Indels 5; Gaps 1;

QY 2 EKTPLT-----TTAAXAPVVXNA 19
| | | | |
Db 21 EEATPLGDGTPTVTAAAPVTVYA 43
```



```

; CURRENT APPLICATION NUMBER: US/10/425,114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 51755
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3067-037-E9_FLI.pep
US-10-425-114A-51755

Query Match      44.4%; Score 36; DB 6; Length 318;
Best Local Similarity 53.8%; Pred. No. 1.7e+02;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      7 LTTAAXAPVVXNA 19
      ||:||:||:||
Db      132 LTSSATAPVLSDA 144

Search completed: January 28, 2004, 13:11:50
Job time : 23 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 13:02:13 ; Search time 41 Seconds
(without alignments)
73.556 Million cell updates/sec

Title: US-09-359-426C-2

Perfect score: 81
Sequence: 1 XEEKTPLTAAAXAPVVXNA 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	76	93.8	19	AAW64480	P. aeruginosa prot
2	76	93.8	19	AAW64480	Pseudomonas aerugi
3	44	54.3	291	AAAG30091	Arabidopsis thalia
4	44	54.3	292	AAAG30090	Arabidopsis thalia
5	44	54.3	306	ABE52865	Escherichia coli p
6	44	54.3	306	AAAB67588	Amino acid sequenc
7	44	54.3	427	AAAG30089	Arabidopsis thalia
8	42	51.9	172	AAAG43604	Arabidopsis thalia
9	42	51.9	215	AAAG43603	Arabidopsis thalia

10	42	51.9	332	21	AAAG3602	Arabidopsis thalia
11	41	50.6	132	24	ABU58158	Rice stress respon
12	41	50.6	1012	22	ABW66926	Drosophila melanog
13	40	49.4	258	23	ABW48482	Listeria monocytog
14	40	49.4	445	22	AAU36108	Klebsiella pneumon
15	40	49.4	498	15	AAAS1691	HIV-type virus WVP
16	40	49.4	498	20	AAW93076	HIV isolate 5180 g
17	40	49.4	500	24	ABJ37450	Benzodiazepines bi
18	40	49.4	788	22	ABW60583	Drosophila melanog
19	40	49.4	1473	22	ABW68024	Drosophila melanog
20	39.5	48.8	1793	22	ABW59613	Drosophila melanog
21	39	48.1	128	21	AAAG3690	Arabidopsis thalia
22	39	48.1	137	22	ABG26339	Novel human diagno
23	39	48.1	159	23	ABJ11251	Yeast selected int
24	39	48.1	159	23	ABG77349	C glutamicum prote
25	39	48.1	233	22	AAAG1909	Selected interacti
26	39	48.1	321	23	AAE22281	Murine SPAS-1 part
27	39	48.1	395	23	AAE22282	Murine tumour SPAS
28	39	48.1	395	23	AAE22283	Murine normal SPAS
29	38.5	47.5	842	22	ABW67966	Drosophila melanog
30	38	46.9	225	22	ABW71511	Drosophila melanog
31	38	46.9	229	22	AAW94775	Human protein sequ
32	38	46.9	229	23	ABW81876	Double strand RNA
33	38	46.9	229	23	AAU81227	Human lung cancer
34	38	46.9	261	22	AAAG1455	C glutamicum prote
35	38	46.9	276	23	ABP73930	Candida albicans e
36	38	46.9	277	22	AAU18229	Novel human DNA-bi
37	38	46.9	277	23	ABG92650	Human DNA-binding
38	38	46.9	307	11	AAW07361	Phospholipase D to
39	38	46.9	521	15	AAW63672	Aldehyde-dehydroge
40	38	46.9	19938	24	ABW98398	Streptomyces virid
41	37	45.7	19	22	AAU27612	Human LEKTI protei
42	37	45.7	87	23	AAO19127	Human circulating
43	37	45.7	120	21	AAAG18092	Arabidopsis thalia
44	37	45.7	243	21	AAW03456	Arabidopsis thalia
45	37	45.7	299	23	ABP66095	Bifidobacterium lo

ALIGNMENTS

RESULT 1
AAW64480
ID AAW64480 standard; peptide; 19 AA.
AC AAW64480;
XX
XX
DT 20-OCT-1998 (first entry)
DE P. aeruginosa protein antigen Pa60 N-terminal peptide fragment.
KW Antigen; Pa60; diagnosis; detection; cystic fibrosis; vaccine;
KW immunogen; infection; treatment.
XX
OS Pseudomonas aeruginosa.
XX
XX
EH Key Location/Qualifiers
FT Misc-difference 1 /label= unknown
FT Misc-difference 12 /label= unknown
FT Misc-difference 17 /label= unknown
FT Misc-difference 17 /label= unknown
XX
PN WO9832769-Al.
XX
XX
PD 30-JUL-1998.
XX
XX
PF 26-JAN-1998; 98WO-CB00217.
XX
XX
PR 24-JAN-1997; 97GB-0001489.
XX
XX
PA (AUSP-) AUSPHARM INT LTD.

PA (CHAP/) CHAPMAN P W.
 XX Clancy RL, Cripps AW, Dunkley M, Kyd J;
 XX WPI; 1998-427879/36.
 XX Protein antigen from *Pseudomonas aeruginosa* and its antigenic
 PT fragments - useful diagnostically to detect specific antibodies,
 PT particularly in patients with cystic fibrosis, and as vaccines
 XX
 PS Claim 3; Page 2; 23pp; English.
 XX
 CC This peptide is the N-terminal fragment of a novel *Pseudomonas aeruginosa*
 CC protein antigen, Pa60. This fragment could be used for diagnostic
 CC detection of *P. aeruginosa* by forming complexes with specific antibodies,
 CC particularly in patients with cystic fibrosis (by analysis of mucus, e.g.
 CC in saliva), or in vaccines or immunogenic compositions to treat or
 CC prevent infection by *P. aeruginosa*.
 XX
 SQ Sequence 19 AA;
 Query Match 93.8%; Score 76; DB 19; Length 19;
 Best Local Similarity 100.0%; Pred. No. 3.6e-06;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 EEKTLTAAAXAPVVXNA 19
 |||||
 DB 2 EEKTLTAAAXAPVVXNA 19
 |||||
 RESULT 2
 AAB69062
 ID AAB69062 standard; peptide; 19 AA.
 XX
 AC AAB69062;
 XX
 DT 18-APR-2001 (first entry)
 XX
 DE *Pseudomonas aeruginosa* protein N-terminal peptide.
 XX
 KW *Pseudomonas aeruginosa*; chitinase; groEL; chiA; antigen; vaccine;
 KW diagnosis; detection; infection; immune response.
 XX
 OS *Pseudomonas aeruginosa*.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1 /note= "unspecified"
 FT Misc-difference 12 /note= "unspecified"
 FT Misc-difference 17 /note= "unspecified"
 FT
 XX WO200102577-A1.
 XX
 XX 11-JAN-2001.
 XX
 XX 03-JUL-2000; 2000WO-GB02554.
 XX
 XX 01-JUL-1999; 99GB-0015419.
 XX
 XX (PROV-) PROVALIS UK LTD.
 XX
 XX Smith CJ, Thompson SE, Smith MW, Peek K, Sizer PJH, Wilkinson MC;
 XX WPI; 2001-080988/09.
 XX
 XX Antigenic *Pseudomonas aeruginosa* proteins, useful in the detection
 PT and/or diagnosis of *P. aeruginosa* infections and for producing vaccines
 PT against *P. aeruginosa* -
 XX
 XX Disclosure; Page 2; 129pp; English.
 XX

CC The present invention describes antigenic *Pseudomonas aeruginosa*
 CC proteins (P1). The *P. aeruginosa* proteins have antibacterial activity
 CC and can be used in vaccines and as antagonists. The proteins or their
 CC fragments, or antibodies are useful in the detection and/or diagnosis
 CC of *P. aeruginosa*. They are also useful for producing a vaccine and
 CC inducing an immune response against *P. aeruginosa* infection. An agent
 CC capable of antagonising, inhibiting or otherwise interfering with the
 CC function or expression of P1 are useful in the manufacture of a
 CC medicament for the treatment or prophylaxis of *P. aeruginosa* infections.
 CC The present sequence represents a probable *P. aeruginosa* protein
 CC N-terminal peptide sequence from the present invention.
 XX
 SQ Sequence 19 AA;
 Query Match 93.8%; Score 76; DB 22; Length 19;
 Best Local Similarity 100.0%; Pred. No. 3.6e-06;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 EEKTLTAAAXAPVVXNA 19
 |||||
 DB 2 EEKTLTAAAXAPVVXNA 19
 |||||
 RESULT 3
 AAG30091
 ID AAG30091 standard; Protein; 291 AA.
 XX
 AC AAG30091;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE *Arabidopsis thaliana* protein fragment SEQ ID NO: 35912.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS *Arabidopsis thaliana*.
 XX
 XX EP1033405-A2.
 XX
 XX 06-SEP-2000.
 XX
 XX 25-FEB-2000; 2000EP-0301439.
 XX
 XX 25-FEB-1999; 99US-0121825.
 XX 05-MAR-1999; 99US-0123180.
 XX 09-MAR-1999; 99US-0123548.
 XX 23-MAR-1999; 99US-0125788.
 XX 25-MAR-1999; 99US-0126264.
 XX 29-MAR-1999; 99US-0126785.
 XX 01-APR-1999; 99US-0127462.
 XX 06-APR-1999; 99US-0128234.
 XX 08-APR-1999; 99US-0128714.
 XX 16-APR-1999; 99US-0129845.
 XX 19-APR-1999; 99US-0130077.
 XX 21-APR-1999; 99US-0130449.
 XX 23-APR-1999; 99US-0130510.
 XX 28-APR-1999; 99US-0130891.
 XX 30-APR-1999; 99US-0132048.
 XX 30-APR-1999; 99US-0132407.
 XX 04-MAY-1999; 99US-0132484.
 XX 05-MAY-1999; 99US-0132485.
 XX 06-MAY-1999; 99US-0132486.
 XX 06-MAY-1999; 99US-0132487.
 XX 07-MAY-1999; 99US-0132863.
 XX 11-MAY-1999; 99US-0134256.
 XX 14-MAY-1999; 99US-0134218.
 XX 14-MAY-1999; 99US-0134219.
 XX 14-MAY-1999; 99US-0134221.
 XX 14-MAY-1999; 99US-0134370.
 XX 18-MAY-1999; 99US-0134768.


```
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 54.3%; Score 44; DB 21; Length 291;
Best Local Similarity 60.0%; Pred. No. 21;
Matches .9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EEKPLTTTAAAPVW 16
   ||| :||| :|||
Db 72 EEKAEMTTAQSPVW 86

RESULT 4
. AAG30090
ID AAG30090 standard; Protein; 292 AA.
XX
AC AAG30090;
XX
DT 17-OCT-2000 (first entry)
XX
DE . Arabidopsis thaliana protein fragment SEQ ID NO: 35911.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121925.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
```

PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 54.3%; Score 44; DB 21; Length 292;
Best Local Similarity 60.0%; Pred. No. 21;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 2 EKTPLTTAAAPVV 16
Db 73 EKAEMTTAMOSPVV 87
RESULT 5
ABBS2865
ID ABBS2865 standard; Protein; 306 AA.
XX
AC ABBS2865;
XX
DT 11-FEB-2002 (first entry)
XX
DE Escherichia coli polypeptide SEQ ID NO 1139.
XX
KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
KW systemic infection; non-diarrhoeal infection; septicemia;
KW pyelonephritis; antibiotic resistance.
XX
OS Escherichia coli.
XX
PN WO200166572-A2.
XX
PD 13-SEP-2001.
XX
PF 12-MAR-2001; 2001WO-EP03445.
XX
PR 10-MAR-2000; 2000PR-0003145.
PR 02-FEB-2001; 2001PR-0001449.
XX
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
XX
DR WPI; 2001-550253/61.
XX
PT A library of DNA fragments of Escherichia coli strains for the
PT phylogenetic determination of a given strain comprises polynucleotides of
PT nature B2/D+ A- -
XX
PS Example 6; Fig 6; 646pp; English.
XX
CC The invention relates to a library of DNA fragments of Escherichia coli
CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
CC and encoded proteins (ABBS2459-ABBS2919 and ABBS2954-ABBS3094) of nature
CC B2/D+A-. The polynucleotides have potential antiinflammatory,
CC antibacterial and immunosuppressive activity as part of pharmaceutical
CC compositions used to treat, palliate or prevent extra-intestinal E. coli
CC infections. The polypeptides are useful for determining the phylogenic
CC group of a given E. coli strain. These polypeptides can detect and treat
CC an undesired development of E. coli, particularly an extra-intestinal
CC infection that include systemic and non-diarrhoeal infections such as
CC septicemia, pyelonephritis and meningitis this is particularly
CC advantageous as bacterial resistance is increasing with the more
CC frequent use of broad spectrum antibiotics.
XX
SQ Sequence 306 AA;

Query Match 54.3%; Score 44; DB 22; Length 306;
Best Local Similarity 47.1%; Pred. No. 22;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 3 EKTPLTTAAAPVVXNA 19
Db 84 EKVPCTSGVAFIVNA 100

RESULT 6

AAB67588
 ID AAB67588 standard; Protein; 306 AA.
 AC AAB67588;
 XX
 XX
 XX 29-MAY-2001 (first entry)
 XX
 DE Amino acid sequence of a deoxyribokinase enzyme.
 XX
 XX Deoxyribonucleoside; deoxyribose 1-phosphate; thymidine phosphorylase;
 KW purine nucleoside phosphorylase; phosphopentose mutase;
 KW phosphopentose aldolase; fructose 1,6-diphosphate aldolase;
 KW deoxyribokinase; nucleoside 2-deoxyribosyltransferase.
 XX
 OS Salmonella typhi.
 XX
 XX WO200114566-A2.
 PN
 XX
 XX 01-MAR-2001.
 PD
 XX 18-AUG-2000; 2000WO-EP08088.
 PF
 XX 20-AUG-1999; 99EP-0116425.
 XX
 XX (HOFF) ROCHE DIAGNOSTICS GMBH.
 PA (INSP) INST PASTEUR.
 PA (PHAR-) PHARMA-WALDHOF GMBH & CO KG.
 XX
 XX Tischer W, Ihlenfeldt H, Barzu O, Sakamoto H, Pistotnik E;
 PI Marliere P, Pochet S;
 PI
 XX WPI; 2001-235026/24.
 DR N-PSDB; AAF55444.
 XX
 XX In vitro enzymatic synthesis of deoxyribonucleosides comprises reacting
 PT deoxyribose 1-phosphate and a nucleobase to form a deoxyribonucleoside
 PT and an inorganic phosphate -
 PT
 XX
 XX Disclosure; Page 59-61; 73pp; English.
 PS
 CC The present sequence represents a deoxyribokinase enzyme. This enzyme
 CC is involved in the biosynthesis of deoxyribonucleosides, and is
 CC used in the method of the invention. The specification describes a
 CC method for the in vitro enzymatic synthesis of deoxyribonucleosides.
 CC The method comprises reacting deoxyribose 1-phosphate and a nucleobase
 CC to form a deoxyribonucleoside and an inorganic phosphate. Enzymes which
 CC may be used in the method of the invention include thymidine
 CC phosphorylase, purine nucleoside phosphorylase, phosphopentose mutase,
 CC phosphopentose aldolase, fructose 1,6-diphosphate aldolase,
 CC deoxyribokinase, and nucleoside 2-deoxyribosyltransferase.
 XX
 SQ Sequence 306 AA;
 Query Match 54.3%; Score 44; DB 22; Length 306;
 Best Local Similarity 47.1%; Pred. No. 22;
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 QY 3 EKTPLTTAAAPVYXNA 19
 |||::|:
 Db 84 EKVPCTSSGVAFVNA 100
 RESULT 7
 AAG30089
 ID AAG30089 standard; Protein; 427 AA.
 XX
 AC AAG30089;
 XX
 XX 17-OCT-2000 (first entry)
 DT
 XX Arabidopsis thaliana protein fragment SEQ ID NO: 35910.
 DE
 XX Protein identification; signal transduction pathway; metabolic pathway;
 XX
 KW

KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX Arabidopsis thaliana.
 OS
 XX EP1033405-A2.
 PN
 XX
 XX 06-SEP-2000.
 PD
 XX
 XX 25-FEB-2000; 2000EP-0301439.
 PF
 XX 25-FEB-1999; 99US-0121825.
 XX 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0136264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.


```
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.

PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 54.3%; Score 44; DB 21; Length 427;
Best Local Similarity 60.0%; Pred.No.32;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 EKTPLTTAAAPVV 16
Db 208 EKAEMTTAMQSPVV 222
||| :||| :|||
||| :||| :|||

RESULT 8
AAG43604
ID AAG43604 standard; Protein; 172 AA.
XX
AC AAG43604;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 54519.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
```

PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132048.
PR 06-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132486.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.

```
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 51.9%; Score 42; DB 21; Length 172;
Best Local Similarity 50.0%; Pred. No. 25;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EKKPLTTAAAXAPVXNA 19
   |||||
Db 131 EKKTPVEKKTGVPVVKKA 148

RESULT 9
AAG43603
ID AAG43603 standard; Protein; 215 AA.
XX
AC AAG43603;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 54518.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-012180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
```

PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 51.9%; Score 42; DB 21; Length 215; Best Local Similarity 50.0%; Pred. No. 32; Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;		
QY	2 BEKPLTTAAXAPVVXNA 19 	
Db	174 EKTPEVKKTGVPVVKKA 191	
RESULT 10		
AAG43602		
ID	AAG43602 standard; Protein; 332 AA.	
XX	AC AAG43602;	
XX	18-OCT-2000 (first entry)	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 54517.	
KW	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
XX	termination sequence.	
OS	Arabidopsis thaliana.	
XX	EP1033405-A2.	
XX	06-SEP-2000.	
XX	25-FEB-2000; 2000EP-0301439.	
PR	25-FEB-1999; 99US-0121825.	
PR	05-MAR-1999; 99US-0123180.	
PR	09-MAR-1999; 99US-0123548.	
PR	23-MAR-1999; 99US-0125788.	
PR	25-MAR-1999; 99US-0126264.	
PR	29-MAR-1999; 99US-0126785.	
PR	01-APR-1999; 99US-0127462.	
PR	06-APR-1999; 99US-0128234.	
PR	08-APR-1999; 99US-0128714.	
PR	16-APR-1999; 99US-0129845.	

PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 51.9%; Score 42; DB 21; Length 332;

Best Local Similarity 50.0%; Pred. No. 52; Indels 0; Gaps 0;

Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 2 EKTPLTTAAAXAPVVXNA 19
 DB 291 EKTPEVKEKTGVVVKKA 308

RESULT 11
 ABUS8158
 ID ABUS8158 standard; Protein; 132 AA.

XX AC ABUS8158;

XX 14-APR-2003 (first entry)

DE Rice stress response protein #4.

XX Plant; EST; expressed sequence tag; stress response; drought;
 KW heat; radiation; pathogen attack; grain flavour; disease resistance;
 KW peptide-methionine sulfoxide reductase; DNA repair; enzyme;
 KW intracellular protein transport.

XX Oryza sativa.

XX US2002152497-A1.

XX 17-OCT-2002.

XX 19-FEB-2002; 2002US-0078929.

XX 07-MAY-1999; 99US-133038P.

XX 07-MAY-1999; 99US-133042P.

XX 11-MAY-1999; 99US-133427P.

XX 11-MAY-1999; 99US-133428P.

XX 11-MAY-1999; 99US-133436P.

XX 11-MAY-1999; 99US-133437P.

XX 11-MAY-1999; 99US-133438P.

XX 04-JUN-1999; 99US-137667P.

XX 05-MAY-2000; 2000US-0566394.

XX (FALC/) FALCO S. C.

XX (FAMO/) FAMODU O. O.

XX (MEYE/) MEYERS B. C.

XX (MIAO/) MIAO G.

XX (ODEL/) ODELL J. T.

XX (RAFA/) RAFALSKI J. A.

PA (THOR/) THORPE C. J.

PA (SAKA/) SAKAI H.

XX (WENG/) WENG Z.

PI Falco SC, Famodu OO, Meyers BC, Miao G, Odell JT, Rafalski JA;

PI Thorpe CJ, Sakai H, Weng Z;

XX WPI; 2003-198391/19.

DR N-PSDB; ABX78329.

XX New peptide-methionine sulfoxide reductase and nucleic acids, useful in

PT improving plant response to stress, engineering plants with increased

PT disease and stress resistance, or and improving/protecting grain flavor

PT

XX Claim 1; Fig 2; 205pp; English.

XX The invention relates to isolated nucleic acids encoding plant stress

CC response proteins (including peptide-methionine sulfoxide

CC reductases) appearing as ABUS8148-ABUS8246 (or a protein 80% identical

CC to them) from Zea mays, Oryza sativa, Glycine max, or Triticum aestivum.

CC Also included are expression cassettes, transformed host cells,

CC transgenic plants/seeds, modulating the level of peptide-methionine

CC sulfoxide reductase in a plant and a computer system/data

CC processing system for identifying, analysing, or modelling a genetic

CC sequence. The plant nucleic acid is useful in developing strategies to

CC improve plant response to stress (e.g. drought, heat, radiation or

CC pathogen attack), engineering plants with increased disease and stress

CC resistance, manipulating DNA repair and recombination efficiency,

CC manipulating intracellular protein transport, and improving/protecting

CC grain flavour. The nucleic acids may also be used as probes or

CC amplification primers in the detection, quantitation or isolation of gene

CC transcripts, for recombinant expression of encoded polypeptides, as

CC immunogens in preparing or screening antibodies, and in sense or

CC antisense suppression of one or more genes in a host cell, tissue or

CC plant. The proteins may be used as immunogens or antigens to obtain

CC antibodies specifically immunoreactive with the protein, and in assays

CC for enzyme agonists or antagonists. The present sequence is a plant

CC stress response protein (or fragment).

XX

XX Query Match 50.6%; Score 41; DB 24; Length 132;

XX Best Local Similarity 50.0%; Pred. No. 28;

XX Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 2 EKTPLTTAAAXAPVVXNA 19

DB 97 EAAPPTTTAAEAAPAIAAA 114

RESULT 12

ABB66926

ID ABB66926 standard; Protein; 1012 AA.

XX AC ABB66926;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 27570.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX This invention describes the isolation of a novel HIV-type retrovirus
 CC called MVP-5180/91 (SCACC V 92092318). Antigens produced from this
 CC product can be used in an assay kit for detecting antibodies against
 CC viruses that cause immune deficiency, preferably where the assay is a
 CC Western blot, ELISA or fluorescence immunoassay. MVP-5180/91, cDNA
 CC and/or antigen can be used for detecting retroviruses that cause immune
 CC deficiency and to prepare vaccines. This sequence represents an HIV
 CC MVP 5180 gag protein.
 CC (Updated on 20-MAR-2003 to correct PF field.)
 CC (Updated on 20-MAR-2003 to correct PR field.)
 XX
 XX Sequence 498 AA;
 SQ

Query Match 49.4%; Score 40; DB 20; Length 498;
 Best Local Similarity 44.4%; Pred. No. 1.8e+02;
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAPVVXNA 19
 DB 119 BETSPRQTSONYPVITNA 136
 ||||| :|: |||||
 ||||| :|: |||||

RESULT 17
 ABJ37450
 ID ABJ37450 standard; Protein; 500 AA.
 AC ABJ37450;
 XX
 DT 08-MAY-2003 (first entry)
 XX
 DE Benzodiazepines biosynthesis protein SEQ ID No 4.
 XX
 KW Benzodiazepine; antrahamycin; biosynthesis.
 XX
 OS Streptomycetes refuineus.
 XX
 PN WO2002101051-A2.
 XX
 PD 19-DEC-2002.
 XX
 PF 11-JUN-2002; 2002WO-CA00864.
 XX
 PR 11-JUN-2001; 2001US-296744P.
 XX
 PA (ECOP-) ECOPIA BIOSCIENCES INC.
 PA (STAF/) STAFFA A.
 PA (FARN/) FARNET C M.
 XX
 PI Staffa A, Farnet CM;
 XX
 DR WPI; 2003-156963/15.
 DR N-PSDB; ABT32131.
 XX
 PT Novel isolated nucleic acid involved in benzodiazepines production,
 PT useful for directly manipulating benzodiazepines and related chemical
 PT structures by chemical engineering of enzymes involved in antrahamycin
 PT biosynthesis -
 XX
 XX Claim 14; Page 113-115; 166pp; English.
 PS
 XX The invention relates to a novel isolated nucleic acid molecule involved
 CC in production of benzodiazepines, in particular antrahamycin. The protein
 CC encoded by the novel nucleic acid of the invention or a protein of 200
 CC consecutive amino acids of that protein are useful for modifying a
 CC biological molecule that is a substrate for a polypeptide encoded by an
 CC antrahamycin gene cluster. The antrahamycin gene cluster present in a
 CC bacterium, preferably Escherichia coli strain DH10B having accession nos.
 CC IDAC 040602-1 and IDAC 040602-2, is useful for producing an antrahamycin
 CC compound or antrahamycin analogue. The genes and proteins provide the
 CC machinery for producing novel compounds based on the structure of
 CC antrahamycins, and allow direct manipulation of antrahamycin and related
 CC chemical structures via chemical engineering of the enzyme involved in

CC the biosynthesis of antrahamycin, modifications which may not be presently
 CC possible by chemical methodology because of complexity of the structures.
 CC The genes and proteins are also useful to introduce chemicals handles
 CC into normally inert positions that permit subsequence chemical
 CC modification, achieving development of novel benzodiazepines, and
 CC generating a focused library of analogues around a benzodiazepine lead
 CC candidate to fine-tune the compound for optimal properties. The nucleic
 CC acids may be used as probes to identify and isolate DNAs encoding
 CC polypeptide sequences, and related nucleic acids. The nucleic acids are
 CC useful for preparing the enzymes involved in the synthesis of
 CC antrahamycins. The polypeptides are useful in a variety of biochemical
 CC reactions and for generating monoclonal or polyclonal antibodies. This
 CC sequence represents a benzodiazepines biosynthesis protein of the
 CC invention.
 XX
 XX Sequence 500 AA;
 SQ

Query Match 49.4%; Score 40; DB 24; Length 500;
 Best Local Similarity 47.1%; Pred. No. 1.8e+02;
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 EKTPLTTAAAXAPVVXNA 19
 DB 198 EOTPLTTTLRLAELTAA 214
 ||||| :|: |||||
 ||||| :|: |||||

RESULT 18
 ABB60583
 ID ABB60583 standard; Protein; 788 AA.
 XX
 AC ABB60583;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 8541.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL04686.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 XX Disclosure; SEQ ID NO 8541; 21pp + Sequence Listing; English.
 PS
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 788 AA;

SQ Query Match 49.4%; Score 40; DB 22; Length 788;

Best Local Similarity 52.9%; Pred. No. 36+02;

Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 EKTPLTTAAAPVXNA 19

Db 602 EDGPTTTAAAPLASAA 618

RESULT 19

ABB68024
ID ABB68024 standard; Protein; 1473 AA.

XX AC ABB68024;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 30864.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL12127.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

PS Disclosure; SEQ ID NO 30864; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1473 AA;

Query Match 49.4%; Score 40; DB 22; Length 1473;

Best Local Similarity 53.3%; Pred. No. 6e+02;

Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAPV 16

Db 1315 DEQAPYTEALGPV 1329

RESULT 20

ABB59613

ID ABB59613 standard; Protein; 1793 AA.

XX AC ABB59613;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 5631.

XX Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL03716.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

PS Disclosure; SEQ ID NO 5631; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1793 AA;

Query Match 48.8%; Score 39.5; DB 22; Length 1793;

Best Local Similarity 52.6%; Pred. No. 9.1e+02;

Matches 10; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

QY 2 EKTPLTT-AAAPVXNA 19

Db 710 KXSSPTTTPARAPVAQNA 728

RESULT 21

AAG39690

ID AAG39690 standard; Protein; 128 AA.

XX AC AAG39690;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 49147.

XX Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

XX	Arabidopsis thaliana.	PR	29-JUN-1999;	99US-0140991.
OS	EP1033405-A2.	PR	30-JUN-1999;	99US-0141287.
XX		PR	01-JUL-1999;	99US-0141842.
PN		PR	01-JUL-1999;	99US-0142154.
XX		PR	02-JUL-1999;	99US-0142055.
PD	06-SEP-2000.	PR	06-JUL-1999;	99US-0142390.
XX		PR	08-JUL-1999;	99US-0142803.
PF	25-FEB-2000; 2000EP-0301439.	PR	09-JUL-1999;	99US-0142920.
XX		PR	12-JUL-1999;	99US-0142977.
XX		PR	13-JUL-1999;	99US-0143542.
PR	25-FEB-1999;	PR	14-JUL-1999;	99US-0143624.
PR	05-MAR-1999;	PR	15-JUL-1999;	99US-0144005.
PR	09-MAR-1999;	PR	16-JUL-1999;	99US-0144085.
PR	23-MAR-1999;	PR	16-JUL-1999;	99US-0144086.
PR	25-MAR-1999;	PR	19-JUL-1999;	99US-0144325.
PR	29-MAR-1999;	PR	19-JUL-1999;	99US-0144331.
PR	01-APR-1999;	PR	19-JUL-1999;	99US-0144332.
PR	06-APR-1999;	PR	19-JUL-1999;	99US-0144333.
PR	08-APR-1999;	PR	19-JUL-1999;	99US-0144334.
PR	16-APR-1999;	PR	19-JUL-1999;	99US-0144335.
PR	19-APR-1999;	PR	20-JUL-1999;	99US-0144352.
PR	21-APR-1999;	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	PR	20-JUL-1999;	99US-0144884.
PR	23-APR-1999;	PR	21-JUL-1999;	99US-0144814.
PR	28-APR-1999;	PR	21-JUL-1999;	99US-0145086.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145088.
PR	30-APR-1999;	PR	22-JUL-1999;	99US-0145087.
PR	04-MAY-1999;	PR	22-JUL-1999;	99US-0145089.
PR	05-MAY-1999;	PR	22-JUL-1999;	99US-0145192.
PR	06-MAY-1999;	PR	23-JUL-1999;	99US-0145145.
PR	07-MAY-1999;	PR	23-JUL-1999;	99US-0145218.
PR	11-MAY-1999;	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145918.
PR	18-MAY-1999;	PR	27-JUL-1999;	99US-0145919.
PR	19-MAY-1999;	PR	28-JUL-1999;	99US-0145951.
PR	20-MAY-1999;	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149902.
PR	21-JUN-1999;	PR	23-AUG-1999;	99US-0149930.
PR	21-JUN-1999;	PR	25-AUG-1999;	99US-0150566.
PR	22-JUN-1999;	PR	26-AUG-1999;	99US-0150884.
PR	23-JUN-1999;	PR	27-AUG-1999;	99US-0151065.
PR	23-JUN-1999;	PR	27-AUG-1999;	99US-0151066.
PR	24-JUN-1999;	PR	27-AUG-1999;	99US-0151080.
PR	24-JUN-1999;	PR	30-AUG-1999;	99US-0151303.
PR	24-JUN-1999;	PR	31-AUG-1999;	99US-0151438.
PR	24-JUN-1999;	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	PR	07-SEP-1999;	99US-0152363.
PR	24-JUN-1999;	PR	10-SEP-1999;	99US-0153070.

PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 48.1%; Score 39; DB 21; Length 128;
 Best Local Similarity 66.7%; Pred. No. 59;
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 EKTPLTTAAXAP 14
 DB 26 EXPSTTTASAP 37.

RESULT 22
 ABC26339
 ID ABG26339 standard; Protein; 137 AA.

XX AC ABG26339;

XX 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #26330.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX N-PSDB; AAS90526.
 DR WPI; 2001-639362/73.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX Claim 20; SEQ ID No 56698; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 137 AA;

Query Match 48.1%; Score 39; DB 22; Length 137;
 Best Local Similarity 63.6%; Pred. No. 63;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 TPLTTAAXAPV 15

DB 104 TPLTTASATPL 114

RESULT 23

ABJ11251
 ID ABJ11251 standard; Protein; 159 AA.

XX AC ABJ11251;

XX 10-DEC-2002 (first entry)

DE Yeast selected interacting domain protein SEQ ID NO: 636.

KW Yeast; protein-protein interaction; Selected Interacting Domain;
 KW SID (RWM); secretion yield; cancer; neurodegenerative disease; fungicide;
 KW cytostatic; neuroprotective.

XX Saccharomyces cerevisiae.

XX WO200266504-A2.

XX 29-AUG-2002.

XX 14-FEB-2002; 2002WO-BF02299.

XX

PR 16-FEB-2001; 2001US-269266P.
 XX (HYBR-) HYBRIGENICS.
 XX Legrain P;
 XX WPI; 2002-674913/72.
 DR N-PSDB; ABT11568.
 XX
 XX New protein-protein complexes of Saccharomyces cerevisiae, useful in
 PT drug screening or development, for developing yeast strains with better
 PT secretion yield of protein, or in gene therapy (e.g. to treat Candida
 PT infection or cancer).
 XX
 PS Claim 6; Page 262; 357pp; English.
 XX
 XX The present invention relates to complexes between Saccharomyces
 CC cerevisiae Selected Interacting Domain (SID (RTM)) proteins and coding
 CC sequences. The protein complexes of S. cerevisiae are useful in drug
 CC development, in screening drugs or agents that modulate the interaction
 CC of proteins, for developing yeast strains with better secretion yield of
 CC protein, and in gene therapy. The protein complexes, polypeptides and
 CC polynucleotides are useful for preventing or treating Candida infection,
 CC cancer or neurodegenerative diseases in humans or animals. The present
 CC sequence is a protein of the invention.
 XX
 SQ Sequence 159 AA; .
 Query Match 48.1%; Score 39; DB 23; Length 159;
 Best Local Similarity 61.5%; Pred. No. 75;
 Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 2 EEKTLTAAAXAP 14
 |||||
 Db 14 EESTPTATAAP 26
 |||||
 RESULT 24
 ABG77349
 ID ABG77349 standard; Protein; 159 AA.
 AC ABG77349;
 XX
 XX 05-NOV-2002 (first entry)
 DT
 DE Selected Interacting Domain (SID) polypeptide #160.
 XX
 XX Yeast; selected interacting domain; SID; antifungal; cancer;
 KW cytostatic; neuroprotective; Candida infection; gene therapy;
 KW neurodegenerative disease.
 XX
 XX Saccharomyces cerevisiae.
 OS
 XX WO200259255-A2.
 PN
 XX
 XX 01-AUG-2002.
 PD
 XX
 XX 25-JAN-2002; 2002WO-EP01350.
 PF
 XX
 XX 26-JAN-2001; 2001US-264577P.
 PR
 XX
 XX (HYBR-) HYBRIGENICS.
 PA
 XX Legrain P;
 XX
 XX WPI; 2002-619165/66.
 DR N-PSDB; ABS62963.
 XX
 XX New complex between two interacting bait and prey Saccharomyces
 PT cerevisiae polypeptides, useful for preventing or treating Candida
 PT infection, cancer or neurodegenerative diseases in a mammal -
 XX
 PS Claim 6; Page 127; 196pp; English.

XX The invention relates to a complex between two interacting Saccharomyces
 CC cerevisiae polypeptides, comprising two Selected Interacting Domain (SID)
 CC polypeptides as bait and prey proteins. A pharmaceutical composition
 CC comprising the complex is useful for preventing or treating Candida
 CC infection, cancer and neurodegenerative diseases in a human or animal,
 CC preferably in a mammal. This sequence represents a SID polypeptide of the
 CC invention.
 XX
 SQ Sequence 159 AA;
 Query Match 48.1%; Score 39; DB 23; Length 159;
 Best Local Similarity 61.5%; Pred. No. 75;
 Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 2 EEKTLTAAAXAP 14
 |||||
 Db 14 EESTPTATAAP 26
 |||||
 RESULT 25
 AAG91909
 ID AAG91909 standard; Protein; 233 AA.
 XX
 AC AAG91909;
 XX
 XX 26-SEP-2001 (first entry)
 DT
 XX
 DE C glutamicum protein fragment SEQ ID NO: 5663.
 XX
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP1108790-A2.
 XX
 XX 20-JUN-2001.
 PD
 XX
 XX 18-DEC-2000; 2000EP-0127688.
 PF
 XX
 XX 16-DEC-1999; 99JP-0377484.
 PR
 XX 07-APR-2000; 2000JP-0159162.
 PR
 XX 03-AUG-2000; 2000JP-0280988.
 PR
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 PA
 XX Nakagawa S, Mizoguchi H, Ando S, Hayaashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 DR WPI; 2001-376931/40.
 DR N-PSDB; AAH67128.
 XX
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX
 PS Claim 17; SEQ ID NO: 5663; 246pp + Sequence Listing; English.
 XX
 XX The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

XX SQ Sequence 233 AA;
 Query Match 48.1%; Score 39; DB 22; Length 233;
 Best Local Similarity 66.7%; Pred. No. 1.1e+02;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 5 TPLTTAAXAPVV 16
 DB 35 TPTTSASAPVV 46
 RESULT 26
 AAE22281
 ID AAE22281 standard; Protein; 321 AA.
 AC AAE22281;
 XX 25-JUL-2002 (first entry)
 XX Murine SPAS-1 partial protein.
 DE XX Murine; protective immunity; cancer; SPAS-1; T cell; pharmaceutical;
 KW prostate; breast; cervix; ovary; placenta; colon; brain; lung; kidney;
 KW chronic lymphocytic leukaemia; germ cell; vaccine; cytostatic; TRAMP;
 KW transgenic adenocarcinoma mouse prostate.
 OS Mus sp.
 XX Key Location/Qualifiers
 FT Misc-difference 155..156 /note= "Encoded by AGCTAAGCA"
 FT Misc-difference 175..176 /note= "Encoded by GGATAAGGG"
 FT Misc-difference 177..178 /note= "Encoded by ACCTAAGCC"
 FT Misc-difference 201..202 /note= "Encoded by TGCTAGTCC"
 FT Misc-difference 204..205 /note= "Encoded by TTCTGACCT"
 FT Misc-difference 224..225 /note= "Encoded by ATTTGACAC"
 FT Misc-difference 239..240 /note= "Encoded by GCATGAAG"
 FT Misc-difference 249..250 /note= "Encoded by GGTGAGGG"
 FT Misc-difference 297..298 /note= "Encoded by TTCTAAGTA"
 FT Misc-difference 300..301 /note= "Encoded by TTCTGACCT"
 FT Misc-difference 310..311 /note= "Encoded by AAATAAATC"
 FT Misc-difference 321 /label= Unknown
 FT /note= "Encoded by AA"
 PN WO200224739-A2.
 XX 28-MAR-2002.
 XX 13-SEP-2001; 2001WO-US28621.
 XX 21-SEP-2000; 2000US-234472P.
 XX (REGC) UNIV CALIFORNIA.
 XX Allison JP, Fasso M, Shastri N;
 XX WPI; 2002-362424/39.
 XX N-PSDB; AAD35418.
 XX New SPAS-1 protein or antigen obtained from TRAMP-C2 tumor cells,
 PT useful as vaccine for treating or inhibiting cancer in patient, e.g.

PT prostate, breast, cervix, ovary, placenta, colon, brain, lung, kidney
 PT or germ cell cancer
 XX Claim 2; Fig 1A; 107pp; English.
 XX The invention relates to compounds and methods for inducing protective
 CC immunity against cancer. The compounds provided include polypeptides
 CC that contain at least one immunogenic portion of one or more SPAS-1
 CC protein and DNA molecules encoding them or antigen obtained from
 CC transgenic adenocarcinoma mouse prostate (TRAMP)-C2 tumour cells. The
 CC immunogenic portion of the SPAS-1 human homologue polynucleotides
 CC sequence, SPAS-1 antibody or its antigen-binding fragment, the antigen-
 CC presenting cell, the T cell population and the pharmaceutical
 CC compositions are useful for inhibiting the development of a cancer in
 CC a patient, specifically prostate, breast, cervix, ovary, placenta,
 CC colon, brain, lung, kidney, chronic lymphocytic leukaemia or germ cell
 CC cancer. In particular, these compounds are useful as vaccines for
 CC inducing protective immunity against cancer. They are also useful for
 CC diagnosing cancer and monitoring cancer progression. The present
 CC sequence is murine SPAS-1 partial protein.
 XX Sequence 321 AA;
 SQ Query Match 48.1%; Score 39; DB 23; Length 321;
 Best Local Similarity 66.7%; Pred. No. 1.6e+02;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 5 TPLTTAAXAPVV 16
 DB 60 SPTTTAATMPVV 71
 RESULT 27
 AAE22282
 ID AAE22282 standard; Protein; 395 AA.
 AC AAE22282;
 XX 25-JUL-2002 (first entry)
 XX Murine tumour SPAS-1 mutant protein.
 DE XX Murine; protective immunity; cancer; SPAS-1; T cell; pharmaceutical;
 KW prostate; breast; cervix; ovary; placenta; colon; brain; lung; kidney;
 KW chronic lymphocytic leukaemia; germ cell; vaccine; cytostatic; TRAMP;
 KW transgenic adenocarcinoma mouse prostate; mutant; mutein.
 XX Mus sp.
 OS Synthetic.
 XX Key Location/Qualifiers
 FT Region 244..252 /note= "Antigenic epitope capable of activating
 FT TRAMP-specific murine T cells"
 FT Misc-difference 251 /note= "Wild-type Arg is substituted with His"
 FT
 XX WO200224739-A2.
 XX 28-MAR-2002.
 XX 13-SEP-2001; 2001WO-US28621.
 XX 21-SEP-2000; 2000US-234472P.
 XX (REGC) UNIV CALIFORNIA.
 XX Allison JP, Fasso M, Shastri N;
 XX WPI; 2002-362424/39.
 XX N-PSDB; AAD35419.
 XX New SPAS-1 protein or antigen obtained from TRAMP-C2 tumor cells,
 PT useful as vaccine for treating or inhibiting cancer in patient, e.g.

PT useful as vaccine for treating or inhibiting cancer in patient, e.g.
PT prostate, breast, cervix, ovary, placenta, colon, brain, lung, kidney
or germ cell cancer -
PS Claim 2; Fig 1D; 107pp; English.
XX
XX The invention relates to compounds and methods for inducing protective
CC immunity against cancer. The compounds provided include polypeptides
CC that contain at least one immunogenic portion of one or more SPAS-1
CC protein and DNA molecules encoding them or antigen obtained from
CC transgenic adenocarcinoma mouse prostate (TRAMP)-C2 tumour cells. The
CC immunogenic portion of the SPAS-1 human homologue polynucleotides
CC sequence, SPAS-1 antibody or its antigen-binding fragment, the antigen-
CC presenting cell, the T cell population and the pharmaceutical
CC compositions are useful for inhibiting the development of a cancer in
CC a patient, specifically prostate, breast, cervix, ovary, placenta,
CC colon, brain, lung, kidney, chronic lymphocytic leukaemia or germ cell
CC cancer. In particular, these compounds are useful as vaccines for
CC inducing protective immunity against cancer. They are also useful for
CC diagnosing cancer and monitoring cancer progression. The present
CC sequence is murine tumour SPAS-1 mutant protein.
XX
SQ Sequence 395 AA;
Query Match 48.1%; Score 39; DB 23; Length 395;
Best Local Similarity 66.7%; Pred. No. 2.1e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 5 TPLTTAAAXAPVV 16
Db :|||||
300 SPTTTAATMPVV 311
RESULT 28
AAE22283
ID AAE22283 standard; Protein; 395 AA.
XX AC AAE22283;
XX
DT 25-JUL-2002 (first entry)
XX DE Murine normal SPAS-1 protein.
XX
KW Murine; protective immunity; cancer; SPAS-1; T cell; pharmaceutical;
KW prostate; breast; cervix; ovary; placenta; colon; brain; lung; kidney;
KW chronic lymphocytic leukaemia; germ cell; vaccine; cytostatic; TRAMP;
KW transgenic adenocarcinoma mouse prostate.
XX
OS Mus sp.
XX
PN WO200224739-A2.
XX
PD 28-MAR-2002.
XX
XX 13-SEP-2001; 2001WO-US28621.
XX
PR 21-SEP-2000; 2000US-234472P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
XX Allison JP, Fasso M, Shastri N;
XX WPI; 2002-362424/39.
DR N-PSDB; AAD35420.
XX
XX New SPAS-1 protein or antigen obtained from TRAMP-C2 tumor cells,
PT useful as vaccine for treating or inhibiting cancer in patient, e.g.
PT prostate, breast, cervix, ovary, placenta, colon, brain, lung, kidney
or germ cell cancer -
XX
PS Claim 2; Fig 1E; 107pp; English.
XX
CC The invention relates to compounds and methods for inducing protective

CC immunity against cancer. The compounds provided include polypeptides
CC that contain at least one immunogenic portion of one or more SPAS-1
CC protein and DNA molecules encoding them or antigen obtained from
CC transgenic adenocarcinoma mouse prostate (TRAMP)-C2 tumour cells. The
CC immunogenic portion of the SPAS-1 human homologue polynucleotides
CC sequence, SPAS-1 antibody or its antigen-binding fragment, the antigen-
CC presenting cell, the T cell population and the pharmaceutical
CC compositions are useful for inhibiting the development of a cancer in
CC a patient, specifically prostate, breast, cervix, ovary, placenta,
CC colon, brain, lung, kidney, chronic lymphocytic leukaemia or germ cell
CC cancer. In particular, these compounds are useful as vaccines for
CC inducing protective immunity against cancer. They are also useful for
CC diagnosing cancer and monitoring cancer progression. The present
CC sequence is murine normal SPAS-1 protein.
XX
SQ Sequence 395 AA;
Query Match 48.1%; Score 39; DB 23; Length 395;
Best Local Similarity 66.7%; Pred. No. 2.1e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 5 TPLTTAAAXAPVV 16
Db :|||||
300 SPTTTAATMPVV 311
RESULT 29
ABB67966
ID ABB67966 standard; Protein; 842 AA.
XX AC ABB67966;
XX
DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 30690.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
DR N-PSDB; ABL12069.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 30690; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
XX
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 842 AA;

Query Match 47.5%; Score 38.5; DB 22; Length 842;
Best Local Similarity 50.0%; Pred. No. 5.8e+02;
Matches 9; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

QY 3 EKTPL-TTAAXAPVXNA 19
||||: |||
Db 598 EKTPIYSTTKAPVWSTS 615

RESULT 30
ABB71511
ID ABB71511 standard; Protein; 225 AA.

XX ABB71511;
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster polypeptide SEQ ID NO 41325.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD

XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
PA
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
DR
XX N-PSDB; ABL15614.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 41325; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 225 AA;
Query Match 46.9%; Score 38; DB 22; Length 225;
Best Local Similarity 58.3%; Pred. No. 1.6e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 TPLTTAAXAPVW 16
|||: |||
Db 201 TPVTPAPTAPVI 212

RESULT 31
AAB94775
ID AAB94775 standard; Protein; 229 AA.

XX AAB94775;
XX 26-JUN-2001 (first entry)
DT
XX Human protein sequence SEQ ID NO:15864.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
KW Homo sapiens.
OS
XX EP1074617-A2.
PN
XX 07-FEB-2001.
PD

XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 15864; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesising polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.

XX Sequence 229 AA;

Query Match 46.9%; Score 38; DB 22; Length 229;
Best Local Similarity 39.3%; Pred. No. 1.7e+02;
Matches 11; Conservative 1; Mismatches 4; Indels 12; Gaps 1;

QY 3 EKTPLTTAA-----XAPVXN 18
|||: |||
Db 109 EKTPLKTTATPLSPKPPRMDTAPVVAS 136


```
RESULT 32
ABB81876
ID ABB81876 standard; Protein; 229 AA.
XX
XX
AC ABB81876;
XX
XX 20-SEP-2002 (first entry)
XX
XX Double strand RNA joint zinc finger protein 25.19.
XX
XX Double strand RNA joint zinc finger protein 25.19; tumour; diabetes;
KW embryo development malformation; immunological disease.
XX
XX Unidentified.
XX
XX CN1339453-A.
XX
XX 13-MAR-2002.
XX
XX 23-AUG-2000; 2000CN-0119719.
XX
XX 23-AUG-2000; 2000CN-0119719.
XX
XX (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX
XX Mao Y, Xie Y;
XX
XX WPI; 2002-455355/49.
XX
XX N-PSDB; ABB86798.
XX
XX New polypeptide-double strand RNA joint zinc finger protein 25.19 and
PT polynucleotide encoding the polypeptide -
XX
XX Claim 1; Page 28 (Disclosure); 34pp; Chinese.
XX
XX The invention relates to a novel polypeptide, double strand RNA joint
CC zinc finger protein 25.19 and the polynucleotide encoding it. The
CC polypeptide is useful in treating various diseases, such as embryo
CC development malformation, tumours, immunological diseases and diabetes.
CC The present invention also discloses the antagonist resisting the
CC polypeptide and its treatment effect, and the application of the
CC polynucleotides for encoding double strand RNA joint zinc finger protein
CC 25.19. The sequence represents the zinc finger protein of the invention.
XX
XX Sequence 229 AA;
SQ
Query Match 46.9%; Score 38; DB 23; Length 229;
Best Local Similarity 39.3%; Pred. No. 1.7e+02;
Matches 11; Conservative 1; Mismatches 4; Indels 12; Gaps 1;
QY 3 EKTPLTTAA-----XAPVYN 18
DB 109 EKTPLKTTATPLSPKPPRMDTAPVVAS 136
RESULT 33
AAU81227
ID AAU81227 standard; Protein; 229 AA.
XX
XX
AC AAU81227;
XX
XX 09-APR-2002 (first entry)
XX
XX Human lung cancer protein, Seq ID No 92.
XX
XX Human; lung cancer; cytostatic; vaccine; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200192525-A2.
XX
XX 06-DEC-2001.
XX
```

```
XX 25-MAY-2001; 2001WO-US17066.
PF
XX
XX 26-MAY-2000; 2000US-207485P.
PR
XX 06-SEP-2000; 2000US-230475P.
PR
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX Harlocker SL, Wang T, Bangur CS, Klee JI, Switzer A;
PI
XX
XX WPI; 2002-122068/16.
XX
XX New lung tumour polypeptides and polynucleotides, useful in
PT pharmaceutical compositions, such as vaccines, for treating or
PT preventing lung cancer, or as probes or primers for nucleic acid
PT hybridisation.
XX
XX Example 8; Page 176; 179pp; English.
XX
XX The invention relates to novel human lung cancer polynucleotide (I)
CC and polypeptides (II). (I) and (II) are useful in pharmaceutical
CC compositions, such as vaccines, for the diagnosis and treatment of lung
CC cancer. The polynucleotides are also useful as probes or primers for
CC nucleic acid hybridisation. ABK24314-ABK24397 represent human lung
CC cancer coding sequences of the invention.
XX
XX Sequence 229 AA;
SQ
Query Match 46.9%; Score 38; DB 23; Length 229;
Best Local Similarity 39.3%; Pred. No. 1.7e+02;
Matches 11; Conservative 1; Mismatches 4; Indels 12; Gaps 1;
QY 3 EKTPLTTAA-----XAPVYN 18
DB 109 EKTPLKTTATPLSPKPPRMDTAPVVAS 136
RESULT 34
AAG91455
ID AAG91455 standard; Protein; 261 AA.
XX
XX AAG91455;
AC
XX 26-SEP-2001 (first entry)
DT
XX
XX C glutamicum protein fragment SEQ ID NO: 5209.
XX
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
XX Corynebacterium glutamicum.
OS
XX
XX EP1108790-A2.
PN
XX
XX 20-JUN-2001.
PD
XX
XX 18-DEC-2000; 2000EP-0127688.
PF
XX
XX 16-DEC-1999; 99JP-0377484.
PR
XX 07-APR-2000; 2000JP-0159162.
PR
XX 03-AUG-2000; 2000JP-0280988.
PR
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
PA
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayaashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI; 2001-376931/40.
DR
XX N-PSDB; AAH66674.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
```


PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226581.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227089.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.

PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-465557/50.
N-PSDB; AAS29105.
Nucleic acid molecules encoding human secreted chromosomal binding proteins, used in preventing, treating or ameliorating a disorder, e.g. Alzheimer's and Parkinson's diseases and cancers -
Claim 11; SEQ ID No 214; 561pp; English.
The present invention relates to the isolation of novel DNA-binding proteins, and cDNA (AAS29030-AAS29157) and genomic sequences encoding for these proteins. DNA-binding proteins such as histones, chromo (chromatin organisation modifier) domain proteins, and Y-box binding proteins may contribute to diseases resulting from aberrant DNA organisation and/or gene transcription. The sequences of the invention are useful in screening assays to identify antagonists and/or agonists that may enhance or block activities mediated by DNA-binding proteins. Blockers of DNA-binding proteins may be useful in treating disorders such as malignant diseases (e.g. cancer), autoimmune disorders

CC (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid
 CC arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious
 CC diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's
 CC disease). The polynucleotide sequences of the invention may also be
 CC used in gene therapy. AAU18154-AAU18281 represent novel DNA-binding
 CC proteins.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 277 AA;

Query Match 46.9%; Score 38; DB 22; Length 277;
 Best Local Similarity 46.7%; Pred. No. 2.1e+02;
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 TPLTTAAAPVXNA 19
 ||: ||: ||:
 Db 120 TPRPAAAPVIVSGA 134

RESULT 37
 ABG92650
 ID ABG92650 standard; Protein; 277 AA.
 XX
 AC ABG92650;
 XX
 DT 18-NOV-2002 (first entry)
 XX
 DE Human DNA-binding protein #76.
 XX
 KW Human; DNA-binding protein; B cell immunodeficiency; autoimmune disorder;
 KW severe combined immunodeficiency; rheumatoid arthritis; Crohn's disease;
 KW diabetes mellitus; allergy; asthma; inflammatory condition; thrombosis;
 KW graft-versus-host disease; blood-related disorder; atherosclerosis;
 KW hyperproliferative disorder; cancer; renal disorder; arrhythmia;
 KW acute glomerulonephritis; cardiovascular disorder; respiratory disorder;
 KW Goodpasture's syndrome; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; endocrine disorder; Addison's disease;
 KW reproductive system disorder; endometriosis; infectious disease;
 KW viral infection; bacterial infection; fungal infection; vaccine;
 KW gastrointestinal disorder; multiple sclerosis; gene therapy.
 XX
 OS Homo sapiens.
 XX
 XX US2002102638-A1.
 XX
 XX 01-AUG-2002.
 XX
 PF 17-JAN-2001; 2001US-0764846.
 XX
 PR 31-JAN-2000; 2000US-179065P.
 PR 04-FEB-2000; 2000US-180628P.
 PR 28-JUN-2000; 2000US-214886P.
 PR 07-JUL-2000; 2000US-216647P.
 PR 07-JUL-2000; 2000US-216880P.
 PR 11-JUL-2000; 2000US-217478P.
 PR 11-JUL-2000; 2000US-217496P.
 PR 14-JUL-2000; 2000US-218290P.
 PR 26-JUL-2000; 2000US-220963P.
 PR 26-JUL-2000; 2000US-220964P.
 PR 14-AUG-2000; 2000US-224518P.
 PR 14-AUG-2000; 2000US-224519P.
 PR 14-AUG-2000; 2000US-225267P.
 PR 14-AUG-2000; 2000US-225268P.
 PR 14-AUG-2000; 2000US-225270P.
 PR 14-AUG-2000; 2000US-225447P.
 PR 14-AUG-2000; 2000US-225757P.
 PR 14-AUG-2000; 2000US-225758P.
 PR 22-AUG-2000; 2000US-226868P.
 PR 30-AUG-2000; 2000US-228924P.
 PR 01-SEP-2000; 2000US-229287P.
 PR 01-SEP-2000; 2000US-229343P.

PR 01-SEP-2000; 2000US-229344P.
 PR 01-SEP-2000; 2000US-229345P.
 PR 05-SEP-2000; 2000US-229509P.
 PR 05-SEP-2000; 2000US-229513P.
 PR 08-SEP-2000; 2000US-231413P.
 PR 21-SEP-2000; 2000US-234223P.
 PR 21-SEP-2000; 2000US-234274P.
 PR 25-SEP-2000; 2000US-234997P.
 PR 27-SEP-2000; 2000US-235834P.
 PR 29-SEP-2000; 2000US-236327P.
 PR 29-SEP-2000; 2000US-236367P.
 PR 29-SEP-2000; 2000US-236368P.
 PR 29-SEP-2000; 2000US-236369P.
 PR 29-SEP-2000; 2000US-236370P.
 PR 02-OCT-2000; 2000US-236802P.
 PR 02-OCT-2000; 2000US-237037P.
 PR 02-OCT-2000; 2000US-237038P.
 PR 02-OCT-2000; 2000US-237039P.
 PR 02-OCT-2000; 2000US-237040P.
 PR 13-OCT-2000; 2000US-239935P.
 PR 20-OCT-2000; 2000US-240960P.
 PR 20-OCT-2000; 2000US-241785P.
 PR 20-OCT-2000; 2000US-241809P.
 PR 01-NOV-2000; 2000US-244617P.
 PR 17-NOV-2000; 2000US-249299P.
 PR 08-DEC-2000; 2000US-251856P.
 PR 08-DEC-2000; 2000US-251868P.
 PR 08-DEC-2000; 2000US-251869P.
 XX
 PA (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX
 PI Rosen CA, Ruben SM, Barash SC;
 XX
 DR WPI; 2002-690611/74.
 DR N-PSDB; ABS68245.
 XX
 PT Novel DNA-binding protein useful for diagnosis, prognosis, prevention
 PT and treatment of immune, hyperproliferative, respiratory,
 PT cardiovascular, reproductive, endocrine, gastrointestinal and
 PT neurological disorders
 XX
 PS Claim 11; SEQ ID No 214; 225pp; English.
 XX
 CC The present invention relates to a new DNA-binding protein. The invention
 CC is useful in treating, preventing, diagnosing and/or prognosing
 CC immunodeficiencies (e.g. B cell immunodeficiencies, severe combined
 CC immunodeficiencies), autoimmune disorders (rheumatoid arthritis, multiple
 CC sclerosis, diabetes mellitus), allergic reactions and conditions (e.g.
 CC asthma), inflammatory conditions, graft-versus-host disease, blood-
 CC related disorders (thrombosis, atherosclerosis), hyperproliferative
 CC disorders (e.g. cancer), renal disorders (e.g. acute glomerulonephritis),
 CC cardiovascular disorders (e.g. arrhythmia), respiratory disorders
 CC (Goodpasture's syndrome), neurological disorders (e.g. Alzheimer's
 CC disease, Parkinson's disease), endocrine disorders (e.g. Addison's
 CC disease), reproductive system disorders (e.g. endometriosis),
 CC infectious diseases (e.g. viral, bacterial or fungal infections) and
 CC gastrointestinal disorders (e.g. Crohn's disease). The invention is also
 CC useful to stimulate neuronal growth and treat, prevent, and/or diagnose
 CC neuronal damage which occurs in certain neuronal disorders or neuro-
 CC degenerative conditions. The present amino acid sequence represents a
 CC human DNA-binding protein of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from USPTO at
 CC http.seqdata.uspto.gov/sequence.
 XX
 XX Sequence 277 AA;

Query Match 46.9%; Score 38; DB 23; Length 277;
 Best Local Similarity 46.7%; Pred. No. 2.1e+02;
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 TPLTTAAAPVVKNA 19
 DE 120 TPRPAAAPVIVSGA 134

RESULT 38
 AAR07361
 ID AAR07361 standard; protein; 307 AA.
 AC AAR07361;
 XX
 DT 25-MAR-2003 (updated)
 DT 31-JAN-1991 (first entry)
 XX
 DE Phospholipase D toxin.
 XX
 KW Phospholipase D toxine gene; vaccine; diagnosis; PLD;
 KW caseous lymphadenitis; CLA.
 XX
 OS Corynebacterium pseudotuberculosis.
 XX
 XX
 FH Key Location/Qualifiers
 FT Peptide 18..47
 FT /label=identified by sequencing purified PLD
 FT
 XX WO9011351-A.
 XX
 PD 04-OCT-1990.
 XX
 PF 29-MAR-1990; 90WO-AU00121.
 XX
 PR 29-MAR-1989; 89AU-0003422.
 XX
 PA (COMW) COMMONWEALTH SERUM LAB COMMISS.
 XX
 PI Nisbet IT, Hodgson ALM, Bird PI, Cox JC, Eggleston DG;
 PI Haynes J;
 XX
 DR WPI: 1990-320256/42.
 DR N-PSDB; AAQ06249.
 XX
 XX Corynebacterium pseudo-tuberculosis phospholipase toxin purificn.
 PT - by concn. crude toxin by ultrafiltration followed by
 PT chromatography on cation exchange resin
 PT
 XX Disclosure; Fig 4(A-C); 58pp; English.
 XX
 CC The PLD gene was expressed in E. coli and the product shown to
 CC have sphingomyelinase activity.
 CC The toxin can be used in the prepn. of vaccines for immunising sheep
 CC against caseous lymphadenitis. The toxin may also be used in
 CC immunoassays for detection of antibody in diagnosis of CLA in sheep
 CC flocks.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 XX
 SQ Sequence 307 AA;
 Query Match 46.9%; Score 38; DB 11; Length 307;
 Best Local Similarity 61.5%; Pred. No. 2.3e+02;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 6 PLTTAAAPVVKXN 18
 DE 18 PVGNAAAAPVVKHN 30

RESULT 39
 AAR63672
 ID AAR63672 standard; Protein; 521 AA.
 AC AAR63672;
 XX
 XX

DT 25-MAR-2003 (updated)
 DT 08-MAY-1995 (first entry)
 XX
 DE Aldehyde-dehydrogenase Aldh-1.
 XX
 KW Retro virus; vector; aldehyde-dehydrogenase;
 KW glutamylcysteine-synthetase; hematopoietic cell; cyclophosphamide;
 KW chemotherapy; transgenic animal; gene therapy; cancer therapy;
 KW selectable marker.
 XX
 OS Homo sapiens.
 XX
 PN WO9423015-A1.
 XX
 PD 13-OCT-1994.
 XX
 PF 01-APR-1994; 94WO-US03624.
 XX
 PR 01-APR-1993; 93US-0041722.
 XX
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 PI Dallafavera R, Gianni A;
 XX
 DR WPI: 1994-333177/41.
 DR N-PSDB; AAQ72450.
 XX
 XX Retroviral vectors encoding human cytosolic aldehyde
 PT dehydrogenase or glutamyl cysteine synthetase - used to transform
 PT a subject's haematopoietic cells to reduce the toxic effects of
 PT cyclo phosphamide chemotherapy
 XX
 PS Disclosure; Fig.4; 92pp; English.
 XX
 CC A novel retro virus vector encodes human cytosolic aldehyde-
 CC dehydrogenase and/or human glutamylcysteine-synthetase (AAR63673).
 CC Hematopoietic cells transfected by the vector are resistant to
 CC cyclophosphamide, providing a means of gene therapy that allows
 CC higher doses of toxic drugs to be used in cancer chemotherapy.
 CC The human genes may also be used as selectable markers for
 CC mammalian cell transfection and for transgenic animal breeding.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC (Updated on 25-MAR-2003 to correct PI field.)
 XX
 SQ Sequence 521 AA;
 Query Match 46.9%; Score 38; DB 15; Length 521;
 Best Local Similarity 44.4%; Pred. No. 4.1e+02;
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAPVVKNA 19
 DE 215 EQTPLTALHVASLKEA 232

RESULT 40
 ABB98398
 ID ABB98398 standard; Protein; 19938 AA.
 XX
 AC ABB98398;
 XX
 DT 05-MAR-2003 (first entry)
 XX
 DE Streptomyces viridochromogenes Avi gene cluster polypeptide frame 1.
 XX
 KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;
 KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.
 XX
 OS Streptomyces viridochromogenes.
 XX
 PN WO200268436-A1.
 XX
 PD 06-SEP-2002.

```

XX 24-AUG-2001; 2001WO-EP09815.
PF
XX
XX 25-FEB-2001; 2001DE-1009166.
PR
XX
XX (COMB-) COMBINATURE BIOPHARM AG.
XX
XX Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;
PI
XX WPI; 2003-018650/01.
XX DR N-PSDB; ABZ37515.
XX
XX New avilamycin derivatives, useful for treatment of infections, and
PT nucleic acid encoding avilamycin synthesis enzymes
XX
XX Example 1; Page 68-301; 319pp; German.
XX
XX The invention relates to avilamycin derivatives (I) with antibacterial,
CC virucide, protozoacide and fungicide activity. (I) are useful for
CC treatment of infections (bacterial, viral, protozoal or fungal), in human
CC or veterinary medicine, particularly where caused by Staphylococcus
CC aureus. (I) are more hydrophilic than known avilamycins. The present
CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces
CC viridochromogenes Avilamycin A biosynthetic gene cluster
CC (ABZ37515-ABZ37516).
XX
XX Sequence 19938 AA;
SQ
Query Match 46.9%; Score 38; DB 24; Length 19938;
Best Local Similarity 58.3%; Pred. No. 2.4e+04;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 TPLTTAAXAPV 16
|| : |||
Db 15995 TPASTRAAAPVI 16006

Search completed: January 28, 2004, 13:08:11
Job time : 42 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 13:07:24 ; Search time 21 Seconds
(without alignments)
38.281 Million cell updates/sec

Title: US-09-359-426C-2

Perfect score: 81
Sequence: 1 XEEKPLTTAAAXAPVVXN 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2.6/prodata/1/iaa/5A_COMB.pep:*
- 2: /cgn2.6/prodata/1/iaa/5B_COMB.pep:*
- 3: /cgn2.6/prodata/1/iaa/6A_COMB.pep:*
- 4: /cgn2.6/prodata/1/iaa/6B_COMB.pep:*
- 5: /cgn2.6/prodata/1/iaa/PCUS_COMB.pep:*
- 6: /cgn2.6/prodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	63	77.8	482	US-09-252-991A-21327	Sequence 21327, A
2	41	50.6	788	US-08-194-338-12	Sequence 12, Appl
3	40	49.4	498	US-08-470-202-59	Sequence 59, Appl
4	40	49.4	498	US-08-471-770-59	Sequence 59, Appl
5	40	49.4	498	US-08-468-059-59	Sequence 59, Appl
6	40	49.4	498	US-09-109-916-59	Sequence 59, Appl
7	40	49.4	498	US-09-886-156-59	Sequence 59, Appl
8	40	49.4	498	US-09-886-149-59	Sequence 59, Appl
9	40	49.4	498	US-09-886-150-59	Sequence 59, Appl
10	40	49.4	498	US-09-886-159-59	Sequence 59, Appl
11	39	48.1	145	US-09-328-352-7139	Sequence 7139, Ap
12	37	45.7	262	US-09-107-532A-5791	Sequence 5791, Ap
13	37	45.7	342	US-09-252-991A-19222	Sequence 19222, A
14	37	45.7	456	US-08-624-125-20	Sequence 20, Appl
15	37	45.7	456	US-08-937-155-20	Sequence 20, Appl
16	37	45.7	463	US-08-679-635A-4	Sequence 4, Appl
17	37	45.7	463	US-09-419-163-4	Sequence 4, Appl
18	36	44.4	19	US-08-943-173-8	Sequence 8, Appl
19	36	44.4	58	US-08-943-173-16	Sequence 16, Appl
20	36	44.4	71	US-08-943-173-2	Sequence 2, Appl
21	36	44.4	162	US-09-252-991A-17602	Sequence 17602, A
22	36	44.4	181	US-09-117-257-19	Sequence 19, Appl
23	36	44.4	181	US-08-945-476-19	Sequence 19, Appl
24	36	44.4	181	US-09-489-352-19	Sequence 19, Appl
25	36	44.4	182	US-09-117-257-48	Sequence 48, Appl
26	36	44.4	182	US-09-489-352-48	Sequence 48, Appl
27	36	44.4	236	US-09-134-001C-3558	Sequence 3558, Ap

28	36	44.4	294	4	US-09-446-301A-6	Sequence 6,
29	36	44.4	295	4	US-09-099-932-6	Sequence 2, Appl
30	36	44.4	355	2	US-08-458-555-2	Sequence 51, Appl
31	36	44.4	507	4	US-09-446-301A-51	Sequence 6, Appl
32	36	44.4	610	1	US-07-821-717B-6	Sequence 6, Appl
33	36	44.4	610	1	US-08-119-262B-6	Sequence 11, Appl
34	36	44.4	610	1	US-08-135-929A-11	Sequence 24, Appl
35	36	44.4	610	1	US-08-234-265A-11	Sequence 24, Appl
36	36	44.4	2318	3	US-09-091-219-24	Sequence 461, App
37	36	44.4	2318	4	US-09-660-541-24	Sequence 26910, A
38	35	43.2	136	4	US-08-936-165A-461	Sequence 4, Appl
39	35	43.2	181	4	US-09-252-991A-26910	Sequence 204, App
40	35	43.2	185	4	US-09-252-991A-32221	Sequence 142, App
41	35	43.2	223	4	US-09-009-816-4	Sequence 137, App
42	35	43.2	225	4	US-09-071-035-204	Sequence 142, App
43	35	43.2	267	3	US-08-818-112-142	Sequence 142, App
44	35	43.2	267	4	US-08-818-111-137	Sequence 142, App
45	35	43.2	267	4	US-09-056-556-142	

ALIGNMENTS

RESULT 1
US-09-252-991A-21327
; Sequence 21327, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21327
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21327

Query Match 77.8%; Score 63; DB 4; Length 482;
Best Local Similarity 82.4%; Pred. No. 0.0075;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 EKKPLTTAAAXAPVVXN 18
|||||
Db 2 EKKPLTTAAAGAPVVDN 18
|||||

RESULT 2
US-08-194-338-12
; Sequence 12, Application US/08194338
; Patent No. 5474898
; GENERAL INFORMATION:
; APPLICANT: Venter, John C.
; APPLICANT: Fraser, Claire M.
; APPLICANT: McCombie, William R.
; TITLE OF INVENTION: OCTOPAMINE RECEPTOR
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson and Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,338
; FILING DATE: 08-FEB-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/676,174
; FILING DATE: 28-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH101.001DV1
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 788 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; US-08-194-338-12

Query Match 50.6%; Score 41; DB 1; Length 788;
Best Local Similarity 52.9%; Pred. No. 64;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 EKTPLTTAAXAPVVXNA 19
Db 602 EDOPTTTAAAPLASAA 618

RESULT 3
US-08-470-202-59
; Sequence 59, Application US/08470202
; Patent No. 5759808
; GENERAL INFORMATION:
; APPLICANT: Guertler, Lutz G.
; APPLICANT: Eberle, Josef
; APPLICANT: Brunn, Albrecht v.
; APPLICANT: Knapp, Stefan
; APPLICANT: Hauser, Hans-Peter
; TITLE OF INVENTION: Retrovirus from the HIV Group and Its
; TITLE OF INVENTION: Use
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,202
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/132,653
; FILING DATE: 05-OCT-1993
; APPLICATION NUMBER: DE P 42 33 646.5
; FILING DATE: 06-OCT-1992

```

```

;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 42 35 718.7
; FILING DATE: 22-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 42 44 541.8
; FILING DATE: 30-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 43 18 186.4
; FILING DATE: 01-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Michael J. Blake
; REGISTRATION NUMBER: 37,096
; REFERENCE/DOCKET NUMBER: 05495-0001-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 498 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-470-202-59

Query Match 49.4%; Score 40; DB 1; Length 498;
Best Local Similarity 44.4%; Pred. No. 56;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 EEKTELTAAAXAPVVXNA 19
Db 119 EETSPTQSONYPIVTNA 136

RESULT 4
US-08-471-770-59
; Sequence 59, Application US/08471770
; Patent No. 5770427
; GENERAL INFORMATION:
; APPLICANT: Guertler, Lutz G.
; APPLICANT: Eberle, Josef
; APPLICANT: Brunn, Albrecht v.
; APPLICANT: Knapp, Stefan
; APPLICANT: Hauser, Hans-Peter
; TITLE OF INVENTION: Retrovirus from the HIV Group and Its
; TITLE OF INVENTION: Use
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,770
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/132,653
; FILING DATE: 05-OCT-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 42 33 646.5
; FILING DATE: 06-OCT-1992
; APPLICATION NUMBER: DE P 42 35 718.7

```


RESULT 7
US-09-886-156-59
; Sequence 59, Application US/09886156
; Patent No. 6528626

Query Match 49.4%; Score 40; DB 4; Length 498;
Best Local Similarity 44.4%; Pred. No. 56;
Matches 8; Conservative 3; Mismatches 7; Indels

```

RESULT 10
US-09-886-159-59
; Sequence 59, Application US/09886159
; Patent No. 6551824
; GENERAL INFORMATION:
; APPLICANT: Guertler, Lutz G.
; APPLICANT: Eberle, Josef
; APPLICANT: Brunn, Albrecht V.
; APPLICANT: Knapp, Stefan
; APPLICANT: Hauser, Hans-Peter
; TITLE OF INVENTION: RETROVIRUS FROM HIV GROUP AND ITS USE
; FILE REFERENCE: 05495.0001-04
; CURRENT APPLICATION NUMBER: US/09/886,159
; CURRENT FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US/09/109,916
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: DE P 42 33 646.5
; PRIOR FILING DATE: 1992-10-06
; PRIOR APPLICATION NUMBER: DE P 42 35 718.7
; PRIOR FILING DATE: 1992-10-22
; PRIOR APPLICATION NUMBER: DE P 42 44 541.8
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: DE P 43 18 186.4
; PRIOR FILING DATE: 1993-06-01
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 59

```

REGISTRATION NUMBER: 40,489

CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/624,125
FILING DATE: 29-MAR-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: KELBER, STEVEN B.
REGISTRATION NUMBER: 30,073
REFERENCE/DOCKET NUMBER: 2747-063-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-624-125-20

Query Match 45.7%; Score 37; DB 1; Length 456;
Best Local Similarity 72.7%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 9 TAAAPVXNA 19
| | | | |
DB 336 TLAAPVANA 346

RESULT 15
US-08-937-155-20
Sequence 20, Application US/08937155
Patent No. 6524811
GENERAL INFORMATION:
APPLICANT: CUNNINGHAM JR., FRANCIS X.
APPLICANT: SUN, ZAIRAN
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND
TITLE OF INVENTION: METABOLISM AND A SYSTEM FOR SCREENING SUCH GENES
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,155
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/624,125
FILING DATE: 29-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: KELBER, STEVEN B.
REGISTRATION NUMBER: 30,073
REFERENCE/DOCKET NUMBER: 2747-063-27
TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-937-155-20

Query Match 45.7%; Score 37; DB 4; Length 456;
Best Local Similarity 72.7%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 9 TAAAPVXNA 19
| | | | |
DB 336 TLAAPVANA 346

RESULT 16
US-08-679-635A-4
Sequence 4, Application US/08679635A
Patent No. 5985643
GENERAL INFORMATION:
APPLICANT: Tomasz, Alexander
APPLICANT: Delencastre, Herminia
TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
STREET: Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/679,635A
FILING DATE: 10-JUL-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-141
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-679-635A-4

Query Match 45.7%; Score 37; DB 2; Length 463;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 8 TTAAPVXNA 19
| | | | |
DB 382 TTAAPLVQTA 393

```
RESULT 17
US-09-419-163-4
; Sequence 4, Application US/09419163
; Patent No. 6391614
; GENERAL INFORMATION:
; APPLICANT: Tomasz, Alexander
; APPLICANT: Delencastre, Herminia
; TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
; TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/419,163
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/679,635
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-141
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 463 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-09-419-163-4
Query Match 45.7%; Score 37; DB 4; Length 463;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 8 TTAAXAPVXNA 19
||| |||
DB 382 TTAAXAPLVQTA 393

RESULT 18
US-08-943-173-8
; Sequence 8, Application US/08943173
; Patent No. 6048538
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; APPLICANT: Shen, Fan
; APPLICANT: Chen, Pei De
; TITLE OF INVENTION: PEPTIDES DERIVED
; TITLE OF INVENTION: FROM THE NON-STRUCTURAL PROTEINS OF
; TITLE OF INVENTION: FOOT AND MOUTH DISEASE VIRUS AS
; TITLE OF INVENTION: DIAGNOSTIC REAGENTS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,173
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lin, Maria C.H.
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4152
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: 212-751-6849
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-943-173-8
Query Match 44.4%; Score 36; DB 3; Length 19;
Best Local Similarity 53.3%; Pred. No. 6.5;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 ERTPLTTAAXAPVV 16
||| |||
DB 5 ERQKPLVKVAKAPVV 19

RESULT 19
US-08-943-173-16
; Sequence 16, Application US/08943173
; Patent No. 6048538
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; APPLICANT: Shen, Fan
; APPLICANT: Chen, Pei De
; TITLE OF INVENTION: PEPTIDES DERIVED
; TITLE OF INVENTION: FROM THE NON-STRUCTURAL PROTEINS OF
; TITLE OF INVENTION: FOOT AND MOUTH DISEASE VIRUS AS
; TITLE OF INVENTION: DIAGNOSTIC REAGENTS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,173
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lin, Maria C.H.
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4152
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: 212-751-6849
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-943-173-8
```

SEQUENCE CHARACTERISTICS:
 LENGTH: 58 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-943-173-16

Query Match 44.4%; Score 36; DB 3; Length 58;
 Best Local Similarity 53.3%; Pred. No. 23;
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAPV 16
 : : : : :
 Db 31 ERQKPLKVKAKAPV 45
 : : : : :
 : : : : :

RESULT 20
 US-08-943-173-2
 ; Sequence 2, Application US/08943173
 ; Patent No. 6048538

GENERAL INFORMATION:
 ; APPLICANT: Wang, Chang Yi
 ; APPLICANT: Shen, Fan
 ; APPLICANT: Chen, Pei De
 ; TITLE OF INVENTION: PEPTIDES DERIVED
 ; FROM THE NON-STRUCTURAL PROTEINS OF
 ; TITLE OF INVENTION: FOOT AND MOUTH DISEASE VIRUS AS
 ; TITLE OF INVENTION: DIAGNOSTIC REAGENTS
 ; NUMBER OF SEQUENCES: 16
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Maria C.H. Lin
 ; STREET: 345 Park Avenue
 ; CITY: New York
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10154-0053

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WORDPERFECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/943,173
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Lin, Maria C.H.
 REGISTRATION NUMBER: 29,323
 REFERENCE/DOCKET NUMBER: 1151-4152
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-758-4800
 TELEFAX: 212-751-6849

INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 71 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-943-173-2

Query Match 44.4%; Score 36; DB 3; Length 71;
 Best Local Similarity 53.3%; Pred. No. 29;
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAPV 16
 : : : : :
 Db 31 ERQKPLKVKAKAPV 45
 : : : : :
 : : : : :

RESULT 21
 US-09-252-991A-17602
 ; Sequence 17602, Application US/09252991A
 ; Patent No. 6551795

GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 17602
 ; LENGTH: 162
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-17602

Query Match 44.4%; Score 36; DB 4; Length 162;
 Best Local Similarity 63.6%; Pred. No. 73;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 KPTLTTAAAXAP 14
 : : : : :
 Db 1 RTPAATAAIAP 11
 : : : : :
 : : : : :

RESULT 22
 US-09-117-257-19
 ; Sequence 19, Application US/09117257
 ; Patent No. 6214355
 ; GENERAL INFORMATION:
 ; APPLICANT: Hook, Magnus
 ; APPLICANT: Guo, Betty
 ; APPLICANT: Hanson, Mark
 ; TITLE OF INVENTION: DbpA AND DbpB COMPOSITIONS AND METHODS OF USE
 ; FILE REFERENCE: 4210.000500
 ; CURRENT APPLICATION NUMBER: US/09/117,257
 ; CURRENT FILING DATE: 1998-07-22
 ; EARLIER APPLICATION NUMBER: PCT/US96/17081
 ; EARLIER FILING DATE: 1996-10-22
 ; EARLIER APPLICATION NUMBER: 08/589,711
 ; EARLIER FILING DATE: 1996-01-22
 ; EARLIER APPLICATION NUMBER: 08/427,023
 ; EARLIER FILING DATE: 1995-04-24
 ; NUMBER OF SEQ ID NOS: 66
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 19
 ; LENGTH: 181
 ; TYPE: PRT
 ; ORGANISM: Borrelia afzelii
 US-09-117-257-19

Query Match 44.4%; Score 36; DB 3; Length 181;
 Best Local Similarity 47.1%; Pred. No. 83;
 Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 3 EKTPLTTAAAXAPVXNA 19
 : : : : :
 Db 139 EKTPTTAAAGIITIAKA 155
 : : : : :
 : : : : :

RESULT 23
 US-08-945-476-19
 ; Sequence 19, Application US/08945476
 ; Patent No. 6248517
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: DECORIN BINDING PROTEIN COMPOSITIONS AND
 ; TITLE OF INVENTION: METHODS OF USE
 ; NUMBER OF SEQUENCES: 27
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICANT: US/08/945,476
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/589,711
FILING DATE: 22-JAN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/427,023
FILING DATE: 24-APR-1995
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-945-476-19

Query Match 44.4%; Score 36; DB 3; Length 181;
Best Local Similarity 47.1%; Pred. No. 83;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 3 EKTPLTTAAXAPVVXNA 19
||||| :
Db 139 EKTPTTTAEGIIITIAKA 155

RESULT 24
US-09-489-352-19
Sequence 19, Application US/09489352
Patent No. 6312907
GENERAL INFORMATION:
APPLICANT: Hook, Magnus
APPLICANT: Guo, Betty
APPLICANT: Hanson, Mark
TITLE OF INVENTION: DbpA AND DbpB COMPOSITIONS AND METHODS OF USE
FILE REFERENCE: 4210.000500
CURRENT APPLICATION NUMBER: US/09/489,352
CURRENT FILING DATE: 2000-01-21
EARLIER APPLICATION NUMBER: PCT/US96/17081
EARLIER FILING DATE: 1996-10-22
EARLIER APPLICATION NUMBER: 08/589,711
EARLIER FILING DATE: 1996-01-22
EARLIER APPLICATION NUMBER: 08/427,023
EARLIER FILING DATE: 1995-04-24
NUMBER OF SEQ ID NOS: 66
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 19
LENGTH: 181
TYPE: PRT
ORGANISM: Borrelia afzelii
US-09-489-352-19

Query Match 44.4%; Score 36; DB 4; Length 181;
Best Local Similarity 47.1%; Pred. No. 83;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 3 EKTPLTTAAXAPVVXNA 19
||||| :
Db 139 EKTPTTTAEGIIITIAKA 155

RESULT 25
US-09-117-257-48
Sequence 48, Application US/09117257
Patent No. 6214355
GENERAL INFORMATION:
APPLICANT: Hook, Magnus
APPLICANT: Guo, Betty
APPLICANT: Hanson, Mark
TITLE OF INVENTION: DbpA AND DbpB COMPOSITIONS AND METHODS OF USE

FILE REFERENCE: 4210.000500
CURRENT APPLICATION NUMBER: US/09/117,257
CURRENT FILING DATE: 1998-07-22
EARLIER APPLICATION NUMBER: PCT/US96/17081
EARLIER FILING DATE: 1996-10-22
EARLIER APPLICATION NUMBER: 08/589,711
EARLIER FILING DATE: 1996-01-22
EARLIER APPLICATION NUMBER: 08/427,023
EARLIER FILING DATE: 1995-04-24
NUMBER OF SEQ ID NOS: 66
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 48
LENGTH: 182
TYPE: PRT
ORGANISM: Borrelia afzelii
US-09-117-257-48

Query Match 44.4%; Score 36; DB 3; Length 182;
Best Local Similarity 47.1%; Pred. No. 84;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 3 EKTPLTTAAXAPVVXNA 19
||||| :
Db 139 EKTPTTTAEGIIITIAKA 155

RESULT 26
US-09-489-352-48
Sequence 48, Application US/09489352
Patent No. 6312907
GENERAL INFORMATION:
APPLICANT: Hook, Magnus
APPLICANT: Guo, Betty
APPLICANT: Hanson, Mark
TITLE OF INVENTION: DbpA AND DbpB COMPOSITIONS AND METHODS OF USE
FILE REFERENCE: 4210.000500
CURRENT APPLICATION NUMBER: US/09/489,352
CURRENT FILING DATE: 2000-01-21
EARLIER APPLICATION NUMBER: PCT/US96/17081
EARLIER FILING DATE: 1996-10-22
EARLIER APPLICATION NUMBER: 08/589,711
EARLIER FILING DATE: 1996-01-22
EARLIER APPLICATION NUMBER: 08/427,023
EARLIER FILING DATE: 1995-04-24
NUMBER OF SEQ ID NOS: 66
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 48
LENGTH: 182
TYPE: PRT
ORGANISM: Borrelia afzelii
US-09-489-352-48

Query Match 44.4%; Score 36; DB 4; Length 182;
Best Local Similarity 47.1%; Pred. No. 84;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 3 EKTPLTTAAXAPVVXNA 19
||||| :
Db 139 EKTPTTTAEGIIITIAKA 155

RESULT 27
US-09-134-001C-3558
Sequence 3558, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964

;; PRIOR FILING DATE: 1997-11-08
;; PRIOR APPLICATION NUMBER: US 60/055,779
;; PRIOR FILING DATE: 1997-08-14
;; NUMBER OF SEQ ID NOS: 5674
;; SEQ ID NO 3558
;; LENGTH: 236
;; TYPE: PRT
;; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3558

Query Match 44.4%; Score 36; DB 4; Length 236;
Best Local Similarity 46.7%; Pred. No. 1.1e+02;
Matches 7; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 4 KXTPLTTAAAXAPVXN 18
|: ||| |||
DB 65 KAPIQTLVALPVVEN 79

RESULT 28
US-09-446-301A-6
; Sequence 6, Application US/09446301A
; Patent No. 6506893
; GENERAL INFORMATION:
; APPLICANT: EL SOLH, NEVINE
; TITLE OF INVENTION: POLYNUCLEOTIDES AND THEIR USE FOR DETECTING RESISTANCE
; TITLE OF INVENTION: TO STREPTOGRAMIN A OR TO STREPTOGRAMIN B AND RELATED
; TITLE OF INVENTION: COMPOUNDS
; FILE REFERENCE: 03715-0059
; CURRENT APPLICATION NUMBER: US/09/446,301A
; CURRENT FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Staphylococcus sp.
US-09-446-301A-6

Query Match 44.4%; Score 36; DB 4; Length 294;
Best Local Similarity 61.5%; Pred. No. 1.4e+02;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 EKTPLTTAAAXAPV 15
|: ||| |||
DB 173 EYEPLTNAAPV 185

RESULT 29
US-09-099-932-6
; Sequence 6, Application US/09099932
; Patent No. 6570001
; GENERAL INFORMATION:
; APPLICANT: El Solh, Nevine
; TITLE OF INVENTION: POLYNUCLEOTIDES AND THEIR USE FOR DETECTING RESISTANCE
; TITLE OF INVENTION: TO STREPTOGRAMIN A OR TO STREPTOGRAMIN B AND RELATED
; TITLE OF INVENTION: COMPOUNDS
; FILE REFERENCE: 03495.0173-00000
; CURRENT APPLICATION NUMBER: US/09/099,932
; CURRENT FILING DATE: 1998-06-19
; EARLIER APPLICATION NUMBER: 60/050,380
; EARLIER FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Staphylococcus
US-09-099-932-6

Query Match 44.4%; Score 36; DB 4; Length 295;

Best Local Similarity 61.5%; Pred. No. 1.4e+02;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 3 EKTPLTTAAAXAPV 15
|: ||| |||
DB 174 EYEPLTNAAPV 186

RESULT 30
US-08-458-555-2
; Sequence 2, Application US/08458555
; Patent No. 5840689
; GENERAL INFORMATION:
; APPLICANT: Daniloff, Joanne K
; APPLICANT: McInties, Elmarie
; TITLE OF INVENTION: Method for Stimulating the Regrowth of
; TITLE OF INVENTION: Neurons
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John H. Runnels
; STREET: P. O. Box 2471
; CITY: Baton Rouge
; STATE: LA
; COUNTRY: USA
; ZIP: 70821-2471
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,555
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/146,734
; FILING DATE: 29-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Runnels, John H
; REGISTRATION NUMBER: 33451
; REFERENCE/DOCKET NUMBER: Attorney No. 5840689 9312
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 504 387-3221
; TELEFAX: 504 346-8049
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-08-458-555-2

Query Match 44.4%; Score 36; DB 2; Length 355;
Best Local Similarity 61.5%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 EKTPLTTAAAXAPV 15
|: ||| |||
DB 308 EYTPLTEPEKAPV 320

RESULT 31
US-09-446-301A-51
; Sequence 51, Application US/09446301A
; Patent No. 6506893
; GENERAL INFORMATION:
; APPLICANT: EL SOLH, NEVINE
; APPLICANT: ALLIGNET, JEANINE

;; TITLE OF INVENTION: POLYNUCLEOTIDES AND THEIR USE FOR DETECTING RESISTANCE
;; TITLE OF INVENTION: TO STREPTOGRAMIN A OR TO STREPTOGRAMIN B AND RELATED
;; FILE OF INVENTION: COMPOUNDS
;; FILE REFERENCE: 03715-0059
;; CURRENT APPLICATION NUMBER: US/09/446,301A
;; CURRENT FILING DATE: 1999-12-20
;; NUMBER OF SEQ ID NOS: 51
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 51
;; LENGTH: 507
;; TYPE: PRT
;; ORGANISM: Staphylococcus sp.
US-09-446-301A-51

Query Match 44.4%; Score 36; DB 4; Length 507;
Best Local Similarity 61.5%; Pred. No. 2.7e+02;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 3 EKTPLTTAAXAPV 15
|: || || || ||
DB 174 EEPPLPTNAAAPV 186

RESULT 32
US-07-821-717B-6
; Sequence 6, Application US/07821717B
; Patent No. 5298239
; GENERAL INFORMATION:
; APPLICANT: Miller, Jonathan L.
; APPLICANT: Cunningham, David
; APPLICANT: Lyle, Vicki A.
; APPLICANT: Finch, Clara N.
; TITLE OF INVENTION: MUTATIONS RENDERING PLATELET
; TITLE OF INVENTION: GLYCOPROTEIN Ib ALPHA LESS REACTIVE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 15-JAN-1992
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Timain, Susan J.
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 20884/21
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1636
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 610 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
PUBLICATION INFORMATION:
AUTHORS: Lopez, Jose A.
AUTHORS: Chung, Dominic W.
AUTHORS: Fujikawa, Kazuo
AUTHORS: Hagen, Frederick S.
AUTHORS: Papayannopoulou, Thalia
AUTHORS: Roth, Gerald J.
TITLE: Cloning of the alpha chain of human

;; TITLE: platelet glycoprotein Ib: A transmembrane protein with homology
;; TITLE: to leucine-rich alpha-2-glycoprotein
;; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
;; VOLUME: 84
;; PAGES: 5615-5619
;; DATE: AUG-1987
;; RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 0 TO 610
;; PUBLICATION INFORMATION:
;; AUTHORS: Zimmerman, Theodore S.
;; AUTHORS: Ruggeri, Zaverio M.
;; AUTHORS: Houghten, Richard A.
;; AUTHORS: Vincete, Vincete
;; AUTHORS: Mohri, Hiroshi
;; TITLE: Proteolytic fragments and synthetic
;; TITLE: peptides that block the binding of von Willebrand factor to the
;; TITLE: platelet membrane glycoprotein Ib
;; DOCUMENT NUMBER: EP 0 317 278 A2
;; FILING DATE: 16-NOV-1988
;; PUBLICATION DATE: 24-MAY-1989
;; RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 0 TO 293
US-07-821-717B-6

Query Match 44.4%; Score 36; DB 1; Length 610;
Best Local Similarity 42.9%; Pred. No. 3.3e+02;
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 3 EKTPLTTAAXAPV 16
|: || || || ||
DB 409 EPTPIATTSPTI 422

RESULT 33
US-08-119-262B-6
; Sequence 6, Application US/08119262B
; Patent No. 5492809
; GENERAL INFORMATION:
; APPLICANT: Miller, Jonathan L.
; APPLICANT: Cunningham, David
; APPLICANT: Lyle, Vicki A.
; APPLICANT: Finch, Clara N.
; TITLE OF INVENTION: MUTATIONS RENDERING PLATELET
; TITLE OF INVENTION: GLYCOPROTEIN Ib ALPHA LESS REACTIVE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 09-SEP-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Timain, Susan J.
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 20884/22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1636
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 610 amino acids
TYPE: amino acid

```

; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PUBLICATION INFORMATION:
;   AUTHORS: Lopez, Jose A.
;   AUTHORS: Chung, Dominic W.
;   AUTHORS: Fujikawa, Kazuo
;   AUTHORS: Hagen, Frederick S.
;   AUTHORS: Papayannopoulou, Thalia
;   AUTHORS: Roth, Gerald J.
; TITLE: Cloning of the alpha chain of human platelet
; TITLE: glycoprotein Ib: A transmembrane protein
; TITLE: leucine-rich alpha-2-glycoprotein
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; VOLUME: 84
; PAGES: 5615-5619
; DATE: AUG-1987
; RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 0 TO 610
; PUBLICATION INFORMATION:
;   AUTHORS: Zimmerman, Theodore S.
;   AUTHORS: Ruggeri, Zaverio M.
;   AUTHORS: Houghten, Richard A.
;   AUTHORS: Vincete, Vincete
;   AUTHORS: Mohri, Hiroshi
; TITLE: Proteolytic fragments and synthetic peptides
; TITLE: that block the binding of von Willebrand
; TITLE: membrane glycoprotein Ib
; DOCUMENT NUMBER: EP 0 317 278 A2
; FILING DATE: 16-NOV-1988
; PUBLICATION DATE: 24-MAY-1989
; RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 0 TO 293
; US-08-119-262B-6
;
; Query Match 44.4%; Score 36; DB 1; Length 610;
; Best Local Similarity 42.9%; Pred. No. 3.3e+02;
; Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
;
; QY 3 EKTPLTTAAXAPVV 16
; Db 409 EPTPIATSTPTI 422
;
; RESULT 34
; US-08-135-929A-11
; Sequence 11, Application US/08135929A
; Patent No. 5593959
; GENERAL INFORMATION:
;   APPLICANT: Miller, Jonathan L.
;   APPLICANT: Cunningham, David
;   APPLICANT: Lyle, Vicki A.
;   APPLICANT: Finch, Clara N.
;   APPLICANT: Pincus, Matthew R.
; TITLE OF INVENTION: Mutations in the Gene Encoding the Alpha
; TITLE OF INVENTION: Chain of Platelet Glycoprotein Ib
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Nixon, Hargrave, Devans & Doyle
;   STREET: Clinton Square, P.O. Box 1051
;   CITY: Rochester
;   STATE: New York
;   COUNTRY: USA
;   ZIP: 14603
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/135,929A
;   FILING DATE: 28-APR-1994
;   CLASSIFICATION: 536
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Timian, Susan J.
;   REGISTRATION NUMBER: 34,103
;   REFERENCE/DOCKET NUMBER: 20884/24
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (716) 263-1636
;   TELEFAX: (716) 263-1600
;   TELEX: 978450
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 610 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
; US-08-135-929A-11
;
; Query Match 44.4%; Score 36; DB 1; Length 610;
; Best Local Similarity 42.9%; Pred. No. 3.3e+02;
; Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
;
; QY 3 EKTPLTTAAXAPVV 16
; Db 409 EPTPIATSTPTI 422
;
; RESULT 35
; US-08-234-265A-11
; Sequence 11, Application US/08234265A
; Patent No. 5624817
; GENERAL INFORMATION:
;   APPLICANT: Miller, Jonathan L.
;   APPLICANT: Cunningham, David
;   APPLICANT: Lyle, Vicki A.
;   APPLICANT: Finch, Clara N.
;   APPLICANT: Pincus, Matthew R.
; TITLE OF INVENTION: Mutations in the Gene Encoding the
; TITLE OF INVENTION: Chain of Platelet Glycoprotein Ib
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Nixon, Hargrave, Devans & Doyle
;   STREET: Clinton Square, P.O. Box 1051
;   CITY: Rochester
;   STATE: New York
;   COUNTRY: USA
;   ZIP: 14603
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/234,265A
;   FILING DATE: 28-APR-1994
;   CLASSIFICATION: 536
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Timian, Susan J.
;   REGISTRATION NUMBER: 34,103
;   REFERENCE/DOCKET NUMBER: 20884/24
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (716) 263-1636
;   TELEFAX: (716) 263-1600
;   TELEX: 978450
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 610 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
; US-08-234-265A-11
;
; Query Match 44.4%; Score 36; DB 1; Length 610;
; Best Local Similarity 42.9%; Pred. No. 3.3e+02;
; Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
;
; QY 3 EKTPLTTAAXAPVV 16
; Db 409 EPTPIATSTPTI 422
;
; with homology
; factor to the

```


Search completed: January 28, 2004, 13:10:29
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 13:08:14 ; Search time 33 Seconds
(without alignments)
119.030 Million cell updates/sec

Title: US-09-359-426C-2

Perfect score: 81
Sequence: 1 XEEKTPTLTAAAPVWVXNA 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 777136 seqs, 206736638 residues

Total number of hits satisfying chosen parameters: 777136

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	55.6	478	12	US-10-369-493-13806
2	44	54.3	306	12	US-10-238-075-1139
3	43	53.1	483	15	US-10-156-761-10589
4	41	50.6	132	14	US-10-078-929-22
5	41	50.6	132	14	US-10-078-929-198
6	40	49.4	346	15	US-10-156-761-13589
7	40	49.4	445	9	US-09-815-242-11701
8	40	49.4	498	10	US-09-886-156-59
9	40	49.4	498	10	US-09-886-150-59
10	40	49.4	498	11	US-09-886-149-59
11	40	49.4	498	11	US-09-886-159-59
12	40	49.4	498	12	US-10-326-090-59
13	40	49.4	500	15	US-10-166-087-4
14	39	48.1	155	10	US-09-952-432A-2
15	39	48.1	155	10	US-09-952-432A-16

16	39	48.1	233	10	US-09-738-626-5663	Sequence 5663, Ap
17	39	48.1	395	10	US-09-952-432A-19	Sequence 19, Appl
18	39	48.1	395	10	US-09-952-432A-21	Sequence 21, Appl
19	39	48.1	477	12	US-10-369-493-8852	Sequence 8852, Ap
20	39	48.1	507	12	US-10-369-493-16863	Sequence 16863, A
21	39	48.1	509	15	US-10-156-761-10971	Sequence 10971, A
22	39	48.1	1353	12	US-10-369-493-5890	Sequence 5890, Ap
23	39	48.1	2080	12	US-10-353-690-36	Sequence 36, Appl
24	38	46.9	229	9	US-09-866-562-92	Sequence 92, Appl
25	38	46.9	261	10	US-09-738-626-5209	Sequence 5209, Ap
26	38	46.9	276	12	US-10-032-585-7767	Sequence 7767, Ap
27	38	46.9	277	10	US-09-784-846-214	Sequence 214, App
28	38	46.9	277	15	US-10-091-483-214	Sequence 214, App
29	38	46.9	478	12	US-10-369-493-13813	Sequence 13813, A
30	38	46.9	495	12	US-10-369-493-4429	Sequence 4429, Ap
31	38	46.9	495	12	US-10-369-493-7187	Sequence 7187, Ap
32	38	46.9	1679	12	US-10-369-493-22080	Sequence 22080, A
33	38	46.9	19695	12	US-10-084-846A-3	Sequence 3, Appli
34	37	45.7	19	12	US-10-220-510-13	Sequence 13, Appl
35	37	45.7	445	12	US-10-369-493-19305	Sequence 19305, A
36	37	45.7	456	10	US-09-323-988D-20	Sequence 20, Appl
37	37	45.7	480	12	US-10-369-493-8054	Sequence 8054, Ap
38	37	45.7	498	10	US-09-323-988D-57	Sequence 57, Appl
39	37	45.7	500	10	US-09-323-988D-58	Sequence 58, Appl
40	37	45.7	500	10	US-09-323-988D-59	Sequence 59, Appl
41	37	45.7	603	15	US-10-128-714-3511	Sequence 3511, Ap
42	37	45.7	618	15	US-10-128-714-8511	Sequence 8511, Ap
43	37	45.7	1064	12	US-10-220-510-1	Sequence 1, Appli
44	37	45.7	1209	12	US-10-080-334-129	Sequence 129, App
45	37	45.7	1387	15	US-10-156-761-13000	Sequence 13000, A

ALIGNMENTS

RESULT 1
US-10-369-493-13806
; Sequence 13806, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkie, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13806
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Pseudomonas fluorescens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(478)
; OTHER INFORMATION: unsure at all xaa locations
US-10-369-493-13806

Query Match 55.6%; Score 45; DB 12; Length 478;
Best Local Similarity 83.3%; Pred. No. 17;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 7 LTAAAPVWVXN 18
Db 1 LTAAAPVWVDN 12

RESULT 2

```

US-10-078-075-1139
; Sequence 1139, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I. N. S. E. R. M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated from
; TITLE OF INVENTION: E. coli, and biological uses of these polynucleotides and of their
; FILE REFERENCE: Blandine
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1139
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-238-075-1139
Query Match 54.3%; Score 44; DB 12; Length 306;
Best Local Similarity 47.1%; Pred. No. 15;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 EKETLTTAAAXAPVVXNA 19
|||:::|:|
Db 84 EKVPCTSSGVAPIFVNA 100

RESULT 3
US-10-156-761-10589
; Sequence 10589, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10589
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10589
Query Match 53.1%; Score 43; DB 15; Length 483;
Best Local Similarity 69.2%; Pred. No. 37;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 6 PLTTTAAAXAPVVXN 18
|||||:::|
Db 5 PLTTEAGAPVADN 17

RESULT 4
US-10-078-929-22
; Sequence 22, Application US/10078929
; Publication No. US20020152497A1
; GENERAL INFORMATION:
; APPLICANT: Rafalski, Antoni
; APPLICANT: Miao, Guo-Hua
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Sakai, Hajime
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Odell, Joan T.
; APPLICANT: Meyers, Blake
; APPLICANT: Thorpe, Catherine
; APPLICANT: Weng, Zude
; TITLE OF INVENTION: Nucleic Acid Fragments Encoding Proteins Involved in
; TITLE OF INVENTION: Stress Response
; FILE REFERENCE: BBI357 US NA
; CURRENT APPLICATION NUMBER: US/10/078,929
; CURRENT FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: 09/566,394
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: 60/133038
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: 60/133042
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: 60/133427
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 60/133437
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 60/133428
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 60/133438
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 60/133436
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 60/137667
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 22
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-078-929-22
Query Match 50.6%; Score 41; DB 14; Length 132;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EEKPTLTTAAAXAPVVXNA 19
|||||:::|
Db 97 EAAPTPTTAAEAPAIAAA 114

RESULT 5
US-10-078-929-198
; Sequence 198, Application US/10078929
; Publication No. US20020152497A1
; GENERAL INFORMATION:
; APPLICANT: Rafalski, Antoni
; APPLICANT: Miao, Guo-Hua
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Sakai, Hajime
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Odell, Joan T.
; APPLICANT: Meyers, Blake
; APPLICANT: Thorpe, Catherine
; APPLICANT: Weng, Zude
; TITLE OF INVENTION: Nucleic Acid Fragments Encoding Proteins Involved in
; TITLE OF INVENTION: Stress Response
; FILE REFERENCE: BBI357 US NA
; CURRENT APPLICATION NUMBER: US/10/078,929
; CURRENT FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: 09/566,394
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: 60/133038
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: 60/133042
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: 60/133427
; PRIOR FILING DATE: 1999-05-11

```

; PRIOR APPLICATION NUMBER: 60/133437
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 60/133428
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 60/133438
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 60/133436
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: 60/137667
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 198
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-078-929-198

Query Match 50.6%; Score 41; DB 14; Length 132;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAPVXNA 19
| | | | | : : :
DB 97 EAAPPTTAAAPAIAAA 114

RESULT 6
US-10-156-761-13589
; Sequence 13589, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13589
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13589

Query Match 49.4%; Score 40; DB 15; Length 346;
Best Local Similarity 66.7%; Pred. No. 83;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 TPLTTAAAPVXNA 19
| | | | | : : :
DB 16 TVLTTAAAAVLSA 30

RESULT 7
US-09-815-242-11701
; Sequence 11701, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11701
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-815-242-11701

Query Match 49.4%; Score 40; DB 9; Length 445;
Best Local Similarity 57.1%; Pred. No. 1.1e+02;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAPV 15
| | | | | : : :
DB 382 EOKTELTAADMAAI 395

RESULT 8
US-09-886-156-59
; Sequence 59, Application US/09886156
; Patent No. US20020155428A1
; GENERAL INFORMATION:
; APPLICANT: Guertler, Lutz G.
; APPLICANT: Eberle, Josef
; APPLICANT: Brunn, Albrecht V.
; APPLICANT: Knapp, Stefan
; APPLICANT: Hauser, Hans-Peter
; TITLE OF INVENTION: RETROVIRUS FROM HIV GROUP AND ITS USE
; FILE REFERENCE: 05495.0001-04
; CURRENT APPLICATION NUMBER: US/09/886,156
; CURRENT FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US/09/109,916
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: DE P 42 33 646.5
; PRIOR FILING DATE: 1992-10-06
; PRIOR APPLICATION NUMBER: DE P 42 35 718.7
; PRIOR FILING DATE: 1992-10-22
; PRIOR APPLICATION NUMBER: DE P 42 44 541.8
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: DE P 43 18 186.4
; PRIOR FILING DATE: 1993-06-01
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 59
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus
US-09-886-156-59

Query Match 49.4%; Score 40; DB 10; Length 498;
Best Local Similarity 44.4%; Pred. No. 1.2e+02;

Matches	8:	Conservative	3:	Mismatches	7:	Indels	0:	Gaps	0:
---------	----	--------------	----	------------	----	--------	----	------	----

Qy 2 EEKPLTTAAXAPVXNA 19
||:| |:|
Db 119 EETSPTSONYPVITNA 136

RESULTS

```

US-09-886-150-59
; Sequence 59, Application US/09886150
; Patent No. US20020172939A1
; GENERAL INFORMATION:
; APPLICANT: Guertler, Lutz G.
; APPLICANT: Eberle, Josef
; APPLICANT: Brunn, Albrecht V.
; APPLICANT: Knapp, Stefan
; APPLICANT: Hauser, Hans-Peter
; TITLE OF INVENTION: RETROVIRUS FROM HIV GROUP AND ITS USE
; FILE REFERENCE: 05495.0001-04
; CURRENT APPLICATION NUMBER: US/09/886,150
; CURRENT FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: 09/109,916
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: DE P 42 33 646.5
; PRIOR FILING DATE: 1992-10-06
; PRIOR APPLICATION NUMBER: DE P 42 35 718.7
; PRIOR FILING DATE: 1992-10-22
; PRIOR APPLICATION NUMBER: DE P 42 44 541.8
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: DE P 43 18 186.4
; PRIOR FILING DATE: 1993-06-01
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 59
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus
US-09-886-150-59

```

Query Match 49.4%; Score 40; DB 10; Length 498;
Best Local Similarity 44.4%; Pred. No. 1.2e+02;
Matches 8; Conservative 3; Mismatches 7; Indels

Qy 2 EEKPLTTAAXAPVVXNA 19
 || : | | : | : |
Db 119 EETSPTSONYPVITNA 136

RESULT 10

RESULI 10
US-09-886-149-59
; Sequence 59, Application US/09886149
; Publication No. US20030003442A1
; GENERAL INFORMATION:
; APPLICANT: Guertler, Lutz G.
; APPLICANT: Eberle, Josef
; APPLICANT: Brunn, Albrecht V.
; APPLICANT: Knapp, Stefan
; APPLICANT: Hauser, Hans-Peter
; TITLE OF INVENTION: RETROVIRUS FROM HIV GROUP AND ITS USE
; FILE REFERENCE: 05495.0001-04
; CURRENT APPLICATION NUMBER: US/09/886,149
; CURRENT FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: 09/109,916
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: DE P 42 33 646.5
; PRIOR FILING DATE: 1992-10-06
; PRIOR APPLICATION NUMBER: DE P 42 35 718.7
; PRIOR FILING DATE: 1992-10-22
; PRIOR APPLICATION NUMBER: DE P 42 44 541.8
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: DE P 43 18 186.4
; PRIOR FILING DATE: 1993-06-01
; NUMBER OF SEQ ID NOS: 67

; SOFTWARE: PatentIn Ver. 2.0

; SEO ID NO 59

: LENGTH: 498

LENGTH: FT
TYPE: DPT

TYPE: PRI

; ORGANISM: HU

Query Match

Query Match 49.4%; Score 40; DB 11; Length 498;
Best Local Similarity 44.4%; Pred. No. 1.2e+02;
Matches 8; Conservative 3; Mismatches 7; Indels

2 EEKTPITTAAXAPVXXNA 19

QY Z E E K T P L T I A A X A P V V X N A 19

[illegible]

RESULT 11

```

US-09-886-159-59
; Sequence 59, Application US/09886159
; Publication No. US20030003443A1
; GENERAL INFORMATION:
; APPLICANT: Guertler, Lutz G.
; APPLICANT: Eberle, Josef
; APPLICANT: Brunn, Albrecht V.
; APPLICANT: Knapp, Stefan
; APPLICANT: Hauser, Hans-Peter
; TITLE OF INVENTION: RETROVIRUS FROM HIV GROUP AND ITS USE
; FILE REFERENCE: 05495.0001-04
; CURRENT APPLICATION NUMBER: US/09/886,159
; CURRENT FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US/09/109,916
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: DE P 42 33 646.5
; PRIOR FILING DATE: 1992-10-06
; PRIOR APPLICATION NUMBER: DE P 42 35 718.7
; PRIOR FILING DATE: 1992-10-22
; PRIOR APPLICATION NUMBER: DE P 42 44 541.8
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: DE P 43 18 186.4
; PRIOR FILING DATE: 1993-06-01
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 59
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus
US-09-886-159-59

```

Query Match	49.4%;	Score 40;	DB 11;	Length 498;
Best Local Similarity	44.4%;	Pred. No. 1.2e+02;		
Matches	8: Conservative	3: Mismatches	7: Indels	

Q. 2. Fe^{2+} and Fe^{3+} ions

QY 2 EEKTPLTAAAPVXNA 19

[illegible]

RESULT 12

```

RESULTS 12
US-10-326-090-59
# Sequence 59, Application US/10326090
# Publication No. US20030166915A1
; GENERAL INFORMATION:
; APPLICANT: Guertler, Lutz G.
; APPLICANT: Eberle, Josef
; APPLICANT: Brunn, Albrecht V.
; APPLICANT: Knapp, Stefan
; APPLICANT: Hauser, Hans-Peter
; TITLE OF INVENTION: RETROVIRUS FROM HIV GROUP AND ITS USE
; FILE REFERENCE: 05495.0001-04
; CURRENT APPLICATION NUMBER: US/10/326,090
; CURRENT FILING DATE: 2002-12-23
; PRIOR APPLICATION NUMBER: US/09/109,916
; PRIOR FILING DATE: 1998-07-02
;

```



```
; PRIOR APPLICATION NUMBER: DE P 42 33 646.5
; PRIOR FILING DATE: 1992-10-06
; PRIOR APPLICATION NUMBER: DE P 42 35 718.7
; PRIOR FILING DATE: 1992-10-22
; PRIOR APPLICATION NUMBER: DE P 42 44 541.8
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: DE P 43 18 186.4
; PRIOR FILING DATE: 1993-06-01
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 59
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus
US-10-326-090-59

Query Match          49.4%; Score 40; DB 12; Length 498;
Best Local Similarity 44.4%; Pred. No. 1.2e+02;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY      2 EKKTLTAAAXAPVXNA 19
Db      119 EETSPRQTSQNPVIVNA 136

RESULT 13
US-10-166-087-4
; Sequence 4, Application US/10166087
; Publication No. US2003007767A1
; GENERAL INFORMATION:
; APPLICANT: Ecopia Biosciences Inc.
; APPLICANT: Farnet, Chris
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: Genes and proteins for the biosynthesis of anthracyclin
; FILE REFERENCE: 3014-20S
; CURRENT APPLICATION NUMBER: US/10/166,087
; CURRENT FILING DATE: 2002-06-11
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Streptomyces rufineus subspecies thermotolerans
US-10-166-087-4

Query Match          49.4%; Score 40; DB 15; Length 500;
Best Local Similarity 47.1%; Pred. No. 1.3e+02;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY      3 EKTPLTAAAXAPVXNA 19
Db      198 EQPLTLRLAEIAEA 214

RESULT 14
US-09-952-432A-2
; Sequence 2, Application US/09952432A
; Patent No. US20020150588A1
; GENERAL INFORMATION:
; APPLICANT: Allison, James P.
; APPLICANT: Fasso, Marcella
; APPLICANT: Shastri, Nilabh
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: SPAS-1 CANCER ANTIGEN
; FILE REFERENCE: 018941-001100S
; CURRENT APPLICATION NUMBER: US/09/952,432A
; CURRENT FILING DATE: 2002-04-15
; PRIOR APPLICATION NUMBER: 60/234,472
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
```

```
; LENGTH: 155
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SPAS-1 CDNA
US-09-952-432A-2

Query Match          48.1%; Score 39; DB 10; Length 155;
Best Local Similarity 66.7%; Pred. No. 50;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      5 TPLTTAAAXAPVV 16
Db      60 SPTTTAATMPVV 71

RESULT 15
US-09-952-432A-16
; Sequence 16, Application US/09952432A
; Patent No. US20020150588A1
; GENERAL INFORMATION:
; APPLICANT: Allison, James P.
; APPLICANT: Fasso, Marcella
; APPLICANT: Shastri, Nilabh
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: SPAS-1 CANCER ANTIGEN
; FILE REFERENCE: 018941-001100S
; CURRENT APPLICATION NUMBER: US/09/952,432A
; CURRENT FILING DATE: 2002-04-15
; PRIOR APPLICATION NUMBER: 60/234,472
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 155
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Predicted AA
; OTHER INFORMATION: sequence of mouse SPAS-1
US-09-952-432A-16

Query Match          48.1%; Score 39; DB 10; Length 155;
Best Local Similarity 66.7%; Pred. No. 50;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      5 TPLTTAAAXAPVV 16
Db      60 SPTTTAATMPVV 71

RESULT 16
US-09-738-626-5663
; Sequence 5663, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
```

```
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patentin ver. 3.0
; SEQ ID NO 5663
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-5663

Query Match      48.1%; Score 39; DB 10; Length 233;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      5 TPLTTAAXAPV 16
DB      35 TPTTSASPAPV 46

RESULT 17
US-09-952-432A-19
; Sequence 19, Application US/09952432A
; Patent No. US20020150588A1
; GENERAL INFORMATION:
; APPLICANT: Allison, James P.
; APPLICANT: Fasso, Marcella
; APPLICANT: Shastri, Nilabh
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: SPAS-1 CANCER ANTIGEN
; FILE REFERENCE: 018941-001110US
; CURRENT APPLICATION NUMBER: US/09/952,432A
; CURRENT FILING DATE: 2002-04-15
; PRIOR APPLICATION NUMBER: 60/234,472
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 19
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Tumor SPAS-1
US-09-952-432A-19

Query Match      48.1%; Score 39; DB 10; Length 395;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      5 TPLTTAAXAPV 16
DB      300 SPTTTAATMPV 311

RESULT 18
US-09-952-432A-21
; Sequence 21, Application US/09952432A
; Patent No. US20020150588A1
; GENERAL INFORMATION:
; APPLICANT: Allison, James P.
; APPLICANT: Fasso, Marcella
; APPLICANT: Shastri, Nilabh
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: SPAS-1 CANCER ANTIGEN
; FILE REFERENCE: 018941-001110US
; CURRENT APPLICATION NUMBER: US/09/952,432A
; CURRENT FILING DATE: 2002-04-15
; PRIOR APPLICATION NUMBER: 60/234,472
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 21
; LENGTH: 395

Query Match      48.1%; Score 39; DB 10; Length 395;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      5 TPLTTAAXAPV 16
DB      300 SPTTTAATMPV 311

us-09-952-432A-21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: No. US20020150588A1mal SPAS-1'
US-09-952-432A-21

Query Match      48.1%; Score 39; DB 10; Length 395;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      5 TPLTTAAXAPV 16
DB      300 SPTTTAATMPV 311

RESULT 19
US-10-369-493-8852
; Sequence 8852, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 8852
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Ralstonia metallidurans
US-10-369-493-8852

Query Match      48.1%; Score 39; DB 12; Length 477;
Best Local Similarity 75.0%; Pred. No. 1.8e+02;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      7 LTTAAXAPVXN 18
DB      1 LTTAAGAPVDN 12

RESULT 20
US-10-369-493-16863
; Sequence 16863, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 16863
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Caulobacter crescentus
US-10-369-493-16863

Query Match      48.1%; Score 39; DB 12; Length 507;
```



```
US-09-866-562-92
; Sequence 92, Application US/09866562
; Patent No. US20020009758A1
; GENERAL INFORMATION:
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Klee, Jennifer
; APPLICANT: Switzer, Anne
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF LUNG CANCER.
; FILE REFERENCE: 210121.502
; CURRENT APPLICATION NUMBER: US/09/866,562
; CURRENT FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 96
; SEQ ID NO 92
; TYPE: PRT
; LENGTH: 229
; ORGANISM: Homo sapiens
US-09-866-562-92

Query Match          46.9%; Score 38; DB 9; Length 229;
Best Local Similarity 39.3%; Pred. No. 1.2e+02;
Matches 11; Conservative 1; Mismatches 4; Indels 12; Gaps 1;

QY      3  EKTPLTTAA-----XAPVVXN 18
      ||||| |
DB      109 EKTPLKTTATPLSPKPPRMDTAPVVAS 136

RESULT 25
US-09-738-626-5209
; Sequence 5209, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5209
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5209

Query Match          46.9%; Score 38; DB 10; Length 261;
Best Local Similarity 70.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      6  PLTTAAAPV 15
      ||||| :||
DB      149 PLTTAASPV 158

RESULT 26
US-09-764-846-214
; Sequence 214, Application US/09764846
; Patent No. US20020102638A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT212
; CURRENT APPLICATION NUMBER: US/09/764,846
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 214
; LENGTH: 277
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (6)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (254)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (261)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (268)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (275)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (277)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-846-214

US-10-032-585-7767
; Sequence 7767, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7767
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7767

Query Match          46.9%; Score 38; DB 12; Length 276;
Best Local Similarity 46.7%; Pred. No. 1.4e+02;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY      4  KPTLTTAAAXAPVVXN 18
      :||| | | | :|
DB      53 ETPLQTRAASPLTLN 67

RESULT 27
US-09-764-846-214
; Sequence 214, Application US/09764846
; Patent No. US20020102638A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT212
; CURRENT APPLICATION NUMBER: US/09/764,846
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 214
; LENGTH: 277
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (6)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (254)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (261)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (268)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (275)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (277)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-846-214
```



```

; SEQ ID NO 7187
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-369-493-7187

Query Match          46.9%; Score 38; DB 12; Length 495;
Best Local Similarity 47.1%; Pred. No. 2.7e+02;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY      3 EKTPLTTAAAXAPVXNA 19
      ||||| |::|
Db      174 EXTPLTALAVNQILSEA 190

RESULT 32
US-10-369-493-22080
; Sequence 3, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22080
; LENGTH: 1679
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22080

Query Match          46.9%; Score 38; DB 12; Length 1679;
Best Local Similarity 46.7%; Pred. No. 1.1e+03;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY      2 EKTPLTTAAAXAPV 16
      ||: :||:|:|
Db      346 EBEATMTTSAVSPTV 360

RESULT 33
US-10-084-846A-3
; Sequence 3, Application US/10084846A
; Publication No. US20040006036A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREPZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR FILING DATE: 2001-08-24
; PRIOR FILING DATE: 2001-02-25
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 3
; LENGTH: 19695
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 1: amino acid sequence encoded by coding strand 1.

```

```

; OTHER INFORMATION: Start codon: gga, Start position: nucleotide 1.
US-10-084-846A-3

Query Match          46.9%; Score 38; DB 12; Length 19695;
Best Local Similarity 58.3%; Pred. No. 1.7e+04;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      5 TPLTTAAAXAPV 16
      ||:|:|:|:|
Db      15780 TPASTRAAAPVI 15791

RESULT 34
US-10-220-510-13
; Sequence 13, Application US/10220510
; Publication No. US20030190637A1
; GENERAL INFORMATION:
; APPLICANT: Hovnanian, Alain
; APPLICANT: Chavanas, Stephane
; APPLICANT: Cookson, William
; APPLICANT: Moffat, Miriam
; APPLICANT: Walley, Andrew
; TITLE OF INVENTION: SUSCEPTIBILITY GENE FOR NETHERTON'S DISEASE
; FILE REFERENCE: I00317.70008.US
; CURRENT APPLICATION NUMBER: US/10/220,510
; CURRENT FILING DATE: 2003-03-04
; PRIOR FILING DATE: 2000-03-02
; PRIOR FILING DATE: 2000-03-02
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Peptide
US-10-220-510-13

Query Match          45.7%; Score 37; DB 12; Length 19;
Best Local Similarity 61.5%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      2 BEKTLTTAAAXAP 14
      |||||:|
Db      3 ESSTPGTTAASMP 15

RESULT 35
US-10-369-493-19305
; Sequence 19305, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19305
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure

```

! LOCATION: (1)..(445)
! OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-19305

Query Match 45.7%; Score 37; DB 12; Length 445;
Best Local Similarity 46.7%; Pred. No. 3.6e+02;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAPVV 16
|||: |||: |||:
Db 21 EETPIQAAALPILL 35

RESULT 36

US-09-323-998D-20
! Sequence 20, Application US/09323998D
! Patent No. US20020102631A1
! GENERAL INFORMATION:
! APPLICANT: CUNNINGHAM JR., FRANCIS X.
! APPLICANT: SUN, ZAIREN
! TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
! TITLE OF INVENTION: METHODS OF USE THEREOF
! FILE REFERENCE: 108172-09019
! CURRENT APPLICATION NUMBER: US/09/323,998D
! CURRENT FILING DATE: 1999-06-02
! PRIOR APPLICATION NUMBER: 09/088,724
! PRIOR FILING DATE: 1998-06-02
! PRIOR APPLICATION NUMBER: 09/088,725
! PRIOR FILING DATE: 1998-06-02
! PRIOR APPLICATION NUMBER: 08/937,155
! PRIOR FILING DATE: 1997-09-25
! PRIOR APPLICATION NUMBER: 08/624,125
! PRIOR FILING DATE: 1996-03-29
! NUMBER OF SEQ ID NOS: 61
! SOFTWARE: PatentIn Ver. 2.1
! SEQ ID NO 20
! LENGTH: 456
! TYPE: PRT
! ORGANISM: Artificial Sequence
! FEATURE:
! OTHER INFORMATION: Description of Artificial Sequence: Consensus
! OTHER INFORMATION: sequence of four plant B-cyclases
US-09-323-998D-20

Query Match 45.7%; Score 37; DB 10; Length 456;
Best Local Similarity 72.7%; Pred. No. 3.7e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 9 TAAXAPVVXNA 19
|||: |||: |||:
Db 336 TLAAAPVVANA 346

RESULT 37

US-10-369-493-8054
! Sequence 8054, Application US/10369493
! Publication No. US20030233675A1
! GENERAL INFORMATION:
! APPLICANT: Cao, Yongwei
! APPLICANT: Hinkle, Gregory J.
! APPLICANT: Slater, Steven C.
! APPLICANT: Goldman, Barry S.
! APPLICANT: Chen, Xianfeng
! TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
! TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
! FILE REFERENCE: 38-10(52052)B
! CURRENT APPLICATION NUMBER: US/10/369,493
! CURRENT FILING DATE: 2003-02-28
! PRIOR APPLICATION NUMBER: US 60/360,039
! PRIOR FILING DATE: 2002-02-21
! NUMBER OF SEQ ID NOS: 47374
! SEQ ID NO 8054
! LENGTH: 480

! TYPE: PRT
! ORGANISM: Rhodobacter sphaeroides
US-10-369-493-8054

Query Match 45.7%; Score 37; DB 12; Length 480;
Best Local Similarity 47.1%; Pred. No. 3.9e+02;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 3 EKTPLTTAAAXAPVVXNA 19
|||: |||: |||:
Db 162 EKTPLTALASMAIFGSA 178

RESULT 38

US-09-323-998D-57
! Sequence 57, Application US/09323998D
! Patent No. US20020102631A1
! GENERAL INFORMATION:
! APPLICANT: CUNNINGHAM JR., FRANCIS X.
! APPLICANT: SUN, ZAIREN
! TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
! TITLE OF INVENTION: METHODS OF USE THEREOF
! FILE REFERENCE: 108172-09019
! CURRENT APPLICATION NUMBER: US/09/323,998D
! CURRENT FILING DATE: 1999-06-02
! PRIOR APPLICATION NUMBER: 09/088,724
! PRIOR FILING DATE: 1998-06-02
! PRIOR APPLICATION NUMBER: 09/088,725
! PRIOR FILING DATE: 1998-06-02
! PRIOR APPLICATION NUMBER: 08/937,155
! PRIOR FILING DATE: 1997-09-25
! PRIOR APPLICATION NUMBER: 08/624,125
! PRIOR FILING DATE: 1996-03-29
! NUMBER OF SEQ ID NOS: 61
! SOFTWARE: PatentIn Ver. 2.1
! SEQ ID NO 57
! LENGTH: 498
! TYPE: PRT
! ORGANISM: Capsicum sp.
US-09-323-998D-57

Query Match 45.7%; Score 37; DB 10; Length 498;
Best Local Similarity 72.7%; Pred. No. 4.1e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 9 TAAXAPVVXNA 19
|||: |||: |||:
Db 370 TLAAAPVVANA 380

RESULT 39

US-09-323-998D-58
! Sequence 58, Application US/09323998D
! Patent No. US20020102631A1
! GENERAL INFORMATION:
! APPLICANT: CUNNINGHAM JR., FRANCIS X.
! APPLICANT: SUN, ZAIREN
! TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
! TITLE OF INVENTION: METHODS OF USE THEREOF
! FILE REFERENCE: 108172-09019
! CURRENT APPLICATION NUMBER: US/09/323,998D
! CURRENT FILING DATE: 1999-06-02
! PRIOR APPLICATION NUMBER: 09/088,724
! PRIOR FILING DATE: 1998-06-02
! PRIOR APPLICATION NUMBER: 09/088,725
! PRIOR FILING DATE: 1998-06-02
! PRIOR APPLICATION NUMBER: 08/937,155
! PRIOR FILING DATE: 1997-09-25
! PRIOR APPLICATION NUMBER: 08/624,125
! PRIOR FILING DATE: 1996-03-29
! NUMBER OF SEQ ID NOS: 61
! SOFTWARE: PatentIn Ver. 2.1
! SEQ ID NO 58

```

; LENGTH: 500
; TYPE: PRT
; ORGANISM: Lycopersicon esculentum
US-09-323-998D-58

Query Match      45.7%; Score 37; DB 10; Length 500;
Best Local Similarity 72.7%; Pred. No. 4.1e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      9 TAAXAPVVXNA 19
DB      372 TLAAAPVVANA 382

```

```

RESULT 40
US-09-323-998D-59
; Sequence 59, Application US/09323998D
; Patent No. US20020102631A1
; GENERAL INFORMATION:
; APPLICANT: CUNNINGHAM JR., FRANCIS X.
; APPLICANT: SUN, ZAIREN
; TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 108172-09019
; CURRENT APPLICATION NUMBER: US/09/323,998D
; CURRENT FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: 09/088,724
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 09/088,725
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 08/937,155
; PRIOR FILING DATE: 1997-09-25
; PRIOR APPLICATION NUMBER: 08/624,125
; PRIOR FILING DATE: 1996-03-29
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 59
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Nicotiana sp.
US-09-323-998D-59

```

```

Query Match      45.7%; Score 37; DB 10; Length 500;
Best Local Similarity 72.7%; Pred. No. 4.1e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      9 TAAXAPVVXNA 19
DB      372 TLAAAPVVANA 382

```

Search completed: January 28, 2004, 13:11:14
Job time : 33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 13:06:34 ; Search time 21 Seconds
(without alignments)
87.010 Million cell updates/sec

Title: US-09-359-426C-2

Perfect score: 81
Sequence: 1 XEKTPLTTAAXAPVVXN 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	63	77.8	482	B83113	catalase PA4236 [i
2	53	65.4	484	A58663	catalase (EC 1.11.
3	48	59.3	480	AB0148	catalase (EC 1.11.
4	46	56.8	262	T33408	hypothetical prote
5	46	56.8	262	E88400	protein H34124.2 [
6	45	55.6	128	JC1273	ribosomal protein
7	44	54.3	306	AE0963	probable carbohydr
8	44	54.3	427	F85436	hypothetical prote
9	43	53.1	483	S37055	catalase (EC 1.11.
10	42	51.9	105	A44639	catalase (EC 1.11.
11	42	51.9	321	T08462	hypothetical prote
12	42	51.9	436	T46107	hypothetical prote
13	41	50.6	132	T50779	copper chaperone h
14	41	50.6	482	S60757	catalase (EC 1.11.
15	40	49.4	171	T31478	hypothetical prote
16	40	49.4	252	H72469	hypothetical prote
17	40	49.4	258	AD1328	hypothetical prote
18	40	49.4	394	E87611	hypothetical prote
19	40	49.4	788	S05661	muscarinic acetyl
20	40	49.4	1360	T31674	hypothetical prote
21	39.5	48.8	405	JQ2147	OHPI protein - mai
22	39	48.1	486	T10772	2-hydroxymuconic s
23	39	48.1	507	B87400	aldehyde dehydroge
24	39	48.1	527	S46088	hypothetical prote
25	39	48.1	1353	T26301	hypothetical prote
26	38.5	47.5	1658	D75489	hypothetical prote
27	38	46.9	143	S68226	growth-blocking pe
28	38	46.9	307	T40089	cytochrome c1, hem
29	38	46.9	307	A35125	phospholipase D [E

30	38	46.9	348	2	S40750	hypothetical prote
31	38	46.9	474	2	E87650	peptidase, M20/M25
32	38	46.9	532	2	S40983	hypothetical prote
33	38	46.9	573	2	F81313	peptidase (M3 fami
34	38	46.9	684	2	T25603	hypothetical prote
35	38	46.9	695	2	D71283	probable translati
36	38	46.9	747	2	F88561	protein F58A4.11 [
37	38	46.9	1679	2	S48385	hypothetical prote
38	37.5	46.3	200	2	G86194	hypothetical prote
39	37	45.7	71	2	F72332	hypothetical prote
40	37	45.7	116	2	C83492	hypothetical prote
41	37	45.7	328	2	T01225	hypothetical prote
42	37	45.7	347	2	T35518	probable secreted
43	37	45.7	463	2	S72992	probable phosphory
44	37	45.7	488	2	T42038	catalase (EC 1.11.
45	37	45.7	500	2	S72505	lycopene beta-cycl

ALIGNMENTS

RESULT 1

B83113 catalase PA4236 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: B83113

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B. adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, ; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: B83113

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-482 <STO>

A:Cross-References: GB:AE004841; GB:AE004091; NID:g9950451; PIDN:AAG07624.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: kata; PA4236

C:Superfamily: catalase

Query Match 77.8%; Score 63; DB 2; Length 482;

Best Local Similarity 82.4%; Pred. No. 0.0041;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 EKTPLTTAAXAPVVXN 18

Db 2 EKTPLTTAAXAPVVXN 18

RESULT 2

A58663 catalase (EC 1.11.1.6) [validated] - Proteus mirabilis

C:Species: Proteus mirabilis

A:Variety: strain Pr, peroxide resistant

C>Date: 19-Nov-1997 #sequence_revision 21-Nov-1997 #text_change 15-Sep-2000

C:Accession: A58663; B58663

R:Buzy, A.; Bracchi, V.; Sterjiades, R.; Chroboczek, J.; Thibault, P.; Gagnon, J.; Jouv

J. Protein Chem. 14, 59-72, 1995

A:Title: Complete amino acid sequence of Proteus mirabilis PR catalase. Occurrence of a

A:Reference number: A58663; MUID:95305957; PMID:7786407

A:Accession: A58663

A:Molecule type: protein

A:Residues: 1-484 <BUZ1>

A:Accession: B58663

A:Molecule type: DNA

A:Residues: 1-305, 'AE', <BUZ2>

R:Gouet, P.; Jouve, H.M.; Dideberg, O.

submitted to the Brookhaven Protein Data Bank, June 1996

A:Reference number: A67899; PDB:2CAE

A:Contents: annotation; X-ray crystallography, 2.2 angstroms, residues 3-475

R;Gouet, P.; Jouve, H.M.; Hajdu, J.
submitted to the Brookhaven Protein Data Bank, June 1996
A:Reference number: A67900; PDB:2CAF
A:Contents: annotation; X-ray crystallography, 2.7 angstroms, residues 3-475
A:Note: compound I
R;Gouet, P.; Jouve, H.M.; Hajdu, J.
submitted to the Brookhaven Protein Data Bank, June 1996
A:Reference number: A67901; PDB:2CAG
A:Contents: annotation; X-ray crystallography, 2.7 angstroms, residues 3-475
A:Note: compound II, dithiothreitol reduced compound I
R;Gouet, P.; Jouve, H.M.; Dideberg, O.
submitted to the Brookhaven Protein Data Bank, July 1996
A:Reference number: A67902; PDB:2CAH
A:Contents: annotation; X-ray crystallography, 2.7 angstroms, residues 3-475
A:Note: native Fe(III) with NADPH
R;Gouet, P.; Jouve, H.M.; Dideberg, O.
J. Mol. Biol. 249, 933-954, 1995
A:Title: Crystal structure of Proteus mirabilis PR catalase with and without bound NADPH
A:Reference number: A58664; PMID:7791219
A:Contents: annotation; X-ray crystallography, 2.2 angstroms
C:Complex: homotetramer
C:Function:
A:Description: catalyzes the conversion of two of molecules of hydrogen peroxide to two
A:Note: this enzyme has a tightly bound NADPH cofactor
C:Superfamily: catalase
C:Keywords: chromoprotein; heme; homotetramer; iron; metalloprotein; NADP; oxidoreductase
F:53/Modified site: methionine sulfone (Met) #status experimental
F:54,93,127/Active site: His, Ser, Asn #status predicted
F:337/Binding site: heme iron (Tyr) (axial ligand) #status experimental

Query Match 65.4%; Score 53; DB 2; Length 484;
Best Local Similarity 75.0%; Pred. No. 0.22; Mismatches 0; Indels 0; Gaps 0;
Matches 12; Conservative 0

QY 3 EKPTLTTAAAXAPVYXN 18
II IIIII IIIII
DB 2 EKXKLLTTAGAPVVDN 17
II IIIII IIIII

RESULT 3
AB0148
catalase (EC 1.11.1.6) [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
C:Accession: AB0148
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; PMID:21470413; PMID:11586360
A:Accession: AB0148
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-480 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC90045.1; PID:gl5979266; GSPDB:GN00175
C:Genetics:
A:Gene: katA
C:Superfamily: catalase
C:Keywords: oxidoreductase

Query Match 59.3%; Score 48; DB 2; Length 480;
Best Local Similarity 68.8%; Pred. No. 1.6; Mismatches 1; Indels 0; Gaps 0;
Matches 11; Conservative 1

QY 3 EKPTLTTAAAXAPVYXN 18
:| IIIII IIIII
DB 4 KXKGLTTAGAPVVDN 19
:| IIIII IIIII

RESULT 4
T33408
hypothetical protein H34124.2 - Caenorhabditis elegans (fragment)

C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33408
R;Iatrelle, P.; Rameley, P.; O'Brien, D.
submitted to the EMBL Data Library, July 1998
A:Description: The sequence of C. elegans cosmid H34124.
A:Reference number: Z21340
A:Accession: T33408
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-262 <LAT>
A:Cross-references: EMBL:AF078794; PIDN:AAC26925.1; GSPDB:GN00021; CESP:H34124.2
A:Experimental source: strain Bristol N2; clone H34124
C:Genetics:
A:Gene: CESP:H34124.2
A:Map position: 3
A:Note: intron positions not resolved (incomplete sequence)

Query Match 56.8%; Score 46; DB 2; Length 262;
Best Local Similarity 56.2%; Pred. No. 1.9; Mismatches 3; Indels 0; Gaps 0;
Matches 9; Conservative 3

QY 4 KTPLTTAAXAPVYXNA 19
IIIIII:|:|:|
DB 190 KTPLTSGSSARVINNA 205
IIIIII:|:|:|

RESULT 5
E88400
protein H34124.2 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: E88400
R;Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; PMID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_el.
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: E88400
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-262 <STO>
A:Cross-references: GB:chr_III; PIDN:AAC26925.1; PID:g3329614; GSPDB:GN00021; CESP:H341
C:Genetics:
A:Gene: H34124.2
A:Map position: 3

Query Match 56.8%; Score 46; DB 2; Length 262;
Best Local Similarity 56.2%; Pred. No. 1.9; Mismatches 3; Indels 0; Gaps 0;
Matches 9; Conservative 3

QY 4 KTPLTTAAXAPVYXNA 19
IIIIII:|:|:|
DB 190 KTPLTSGSSARVINNA 205
IIIIII:|:|:|

RESULT 6
JC1273
ribosomal protein L7/L12 - Streptomyces antibioticus
C:Species: Streptomyces antibioticus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 13-Aug-1999
C:Accession: JC1273
R;Parra, F.; Blanco, G.; Alonso, J.M.M.; Balbin, M.; Mendez, C.; Salas, J.A.
Gene 118, 127-129, 1992
A:Title: Cloning and sequence of a gene encoding the L7/L12 ribosomal protein equivalent
A:Reference number: JC1273; PMID:92380478; PMID:1511874
A:Accession: JC1273
A:Molecule type: DNA
A:Residues: 1-128 <PAR>
A:Cross-references: GB:M89911; NID:gl53436; PIDN:AAA26811.1; PID:gl53438
C:Superfamily: Escherichia coli ribosomal protein L12
C:Keywords: protein biosynthesis; ribosome

```
Query Match          55.6%; Score 45; DB 2; Length 128;
Best Local Similarity 66.7%; Pred. No. 1.3;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKTPLTTTAAAPV 16
   ||| : |||
Db 28 EKFEDVTAAAPV 42

RESULT 7
AE0963
probable carbohydrate kinase STY3989 [imported] - Salmonella enterica serovar typhi
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AE0963
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AE0963
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-306 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD03201.1; PID:g16504832; GSPDB:GN00176
C:Genetics:
A:Gene: STY3989
C:Superfamily: ribokinase

Query Match          54.3%; Score 44; DB 2; Length 306;
Best Local Similarity 47.1%; Pred. No. 4.9;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 EKTPLTTTAAAPV 19
   ||| : |||
Db 84 EKVPCTSSGVAFVNA 100

RESULT 8
F85436
hypothetical protein AT4g36970 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: F85436
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488; PMID:10617198
A:Accession: F85436
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-427 <STO>
A:Cross-references: GB:NC_001268; NID:g7270646; PIDN:CAB80363.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4g36970
A:Map position: 4

Query Match          54.3%; Score 44; DB 2; Length 427;
Best Local Similarity 60.0%; Pred. No. 7.1;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKTPLTTTAAAPV 16
   ||| : |||
Db 208 EEKAEMTTAQSPV 222

RESULT 9
S37055
catalase (EC 1.11.1.6) alpha-2 chain - Streptomyces violaceus
```

```
C:Species: Streptomyces violaceus
C:Date: 10-Dec-1993 #sequence_revision 26-May-1995 #text_change 04-Mar-2000
C:Accession: S37055
R:Facey, S.; van Pee, K.H.; Vining, L.C.
submitted to the EMBL Data Library, August 1993
A:Reference number: S37055
A:Accession: S37055
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-483 <PAC>
A:Cross-references: EMBL:X74791; NID:g397888; PIDN:CAAS2796.1; PID:g581780
A:Note: the source is designated as Streptomyces venezuelae
C:Genetics:
A:Start codon: GTG
C:Superfamily: catalase
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
F:54,93,127/Active site: His, Ser, Asn #status predicted
F:337/Binding site: heme iron (Tyr) (axial ligand) #status predicted

Query Match          53.1%; Score 43; DB 2; Length 483;
Best Local Similarity 69.2%; Pred. No. 12;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 6 PLTTTAAAPV 18
   ||| : |||
Db 5 PLTTTGAAPVADN 17

RESULT 10
A44639
catalase (EC 1.11.1.6) - Streptomyces coelicolor (fragments)
C:Species: Streptomyces coelicolor
C:Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 31-Oct-1997
C:Accession: A44639
R:Walker, G.
submitted to the Protein Sequence Database, September 1994
A:Reference number: A44639
A:Accession: A44639
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-105 <WAL>
C:Superfamily: catalase
C:Keywords: oxidoreductase

Query Match          51.9%; Score 42; DB 2; Length 105;
Best Local Similarity 69.2%; Pred. No. 3.4;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 6 PLTTTAAAPV 18
   ||| : |||
Db 7 PLTTTGAAPVADN 19

RESULT 11
T08462
hypothetical protein F2206.250 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999
C:Accession: T08462
R:Quetier, F.; Purnelle, B.; Boutry, M.; Goffeau, A.; Salanoubat, M.; Mewes, H.W.; Maye
submitted to the Protein Sequence Database, May 1999
A:Reference number: Z16420
A:Accession: T08462
A:Molecule type: DNA
A:Residues: 1-321 <QUE>
A:Cross-references: EMBL:AL050300; GSPDB:GN00061; ATSP:F2206.250
A:Experimental source: cultivar Columbia; BAC clone F2206
C:Genetics:
A:Gene: ATSP:F2206.250
A:Map position: 3

Query Match          51.9%; Score 42; DB 2; Length 321;
Best Local Similarity 50.0%; Pred. No. 12;
```

Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EEXTPLTTTAAAXAPVXNA 19
|||||: ||| |
Db 291 EEXTPVEKKTGVVVKKA 308

RESULT 12
T46107
hypothetical protein T25B15.140 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46107
R;Alcaraz, J.P.; Clabault, G.; Cottet, A.; Mache, R.; Mewes, H.W.; Lemcke, K.; Mayer, K.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23021
A:Status: preliminary
A:Accession: T46107
A:Molecule type: DNA
A:Residues: 1-436 <ALC>
A:Cross-references: EMBL:AL132972
A:Experimental source: cultivar Columbia; BAC clone T25B15
C:Genetics:
A:Map position: 3
A:Introns: 418/1
A:Note: T25B15.140

Query Match 51.9%; Score 42; DB 2; Length 436;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EEXTPLTTTAAAXAPVXNA 19
|||||: ||| |
Db 395 EEXTPVEKKTGVVVKKA 412

RESULT 13
T50779
copper chaperone homolog CCH [imported] - rice
C:Species: Oryza sativa (rice)
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Dec-2002
C:Accession: T50779
R;Himeiblaue, E.; Mira, H.; Lin, S.J.; Culotta, V.C.; Penarrubia, L.; Amasino, R.M.
Plant Physiol. 117, 1227-1234, 1998
A:Title: Identification of a functional homolog of the yeast copper homeostasis gene ATX
A:Reference number: Z24450; MUID:9701579; PMID:9701579
A:Accession: T50779
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-132 <HIM>
A:Cross-references: EMBL:AF198626; PIDN:AAF15285.1
C:Superfamily: copper homeostasis factor

Query Match 50.6%; Score 41; DB 2; Length 132;
Best Local Similarity 50.0%; Pred. No. 6.5;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EEXTPLTTTAAAXAPVXNA 19
|||||: ||| |
Db 97 EAAPPTTAAEAPAAIAA 114

RESULT 14
S60757
catalase (EC 1.11.1.6) - Bordetella pertussis
C:Species: Bordetella pertussis
C:Date: 27-Apr-1996 #sequence_revision 13-Mar-1997 #text_change 04-Mar-2000
C:Accession: S60757
R;DeShazer, D.; Wood, G.E.; Friedman, R.L.
Mol. Microbiol. 14, 123-130, 1994
A:Title: Molecular characterization of catalase from Bordetella pertussis: identification
A:Reference number: S60757; MUID:95131725; PMID:7830550
A:Accession: S60757

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-482 <DES>
A:Cross-references: EMBL:U07800; NID:G494943; PIDN:AAA18481.1; PID:G494944
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1994
C:Genetics:
A:Gene: catA
C:Superfamily: catalase
C:Keywords: Chromoprotein, heme; iron; metalloprotein; oxidoreductase
F:57,96,130/Active site: His, Ser, Asn #status Predicted
F:340/Binding site: heme iron (Tyr) (axial ligand) #status predicted
Query Match 50.6%; Score 41; DB 2; Length 482;
Best Local Similarity 75.0%; Pred. No. 27;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 LTTTAAAXAPVXN 18
|||||: ||| |
Db 9 LTTTAAAGAPVADN 20

RESULT 15
T31478
hypothetical protein F56F12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T31478
R;Ainscough, R.
submitted to the EMBL Data Library, October 1999
A:Reference number: Z21039
A:Accession: T31478
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-171 <WIL>
A:Cross-references: EMBL:Z82273; PIDN:CAB54978.1; GSPDB:GN00022; CESP:F56F12.1
A:Experimental source: clone F56F12
C:Genetics:
A:Gene: CESP:F56F12.1
A:Map position: 4
A:Introns: 16/2; 50/1

Query Match 49.4%; Score 40; DB 2; Length 171;
Best Local Similarity 53.3%; Pred. No. 13;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 5 TPLTTTAAAXAPVXNA 19
|||||: ||| |
Db 73 TPTTTTAAAPIAAVA 87

RESULT 16
H72469
hypothetical protein APE2401 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: H72469
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takai
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropy
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: H72469
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-252 <KAW>
A:Cross-references: DDBJ:AP000064; NID:G5105945; PIDN:BAA81416.1; PID:d1045202; PID:g51
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE2401

Query Match 49.4%; Score 40; DB 2; Length 252;
Best Local Similarity 88.9%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 6 PLTTAAXAP 14
Db 64 PLTTAASAP 72

RESULT 17

AD1328
hypothetical protein lmo2028 [imported] - *Listeria monocytogenes* (strain EGD-e)
C/Species: *Listeria monocytogenes*
C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C/Accession: AD1328
R/Glasz, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.N.; Karst, U.
Science 294, 849-852, 2001
A/Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A.; Title: Comparative genomics of *Listeria* species.
A/Reference number: AB1077; MUID:21537279; PMID:11679669
A/Accession: AD1328
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-258 <GLA>
A/Cross-references: GB:NC_003210; PIDN:CAD00106.1; PID:gl6411499; GSPDB:GN00177
A/Experimental source: strain EGD-e
C/Genetics:
A/Gene: lmo2028

Query Match 49.4%; Score 40; DB 2; Length 258;
Best Local Similarity 57.1%;
Pred. No. 20;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 3 EKPTLTAAAPV 16
 | : | | | |
Db 158 EETELTNVYAPVI 171

RESULT 18

E87611

hypothetical protein CC2927 [imported] - Caulobacter crescentus

C;Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C;Accession: E87611

C;Author: R.Nielsen, W.C. Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonin, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of *Caulobacter crescentus*.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: E87611

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-394 <STO>

A;Cross-references: GB:AE005673; NID:g13424551; PIDN:AAK24889.1; GSPDB:GN00148

C;Genetics:

A;Gene: CC2927

Query Match	49.4%	Score 40;	DB 2;	Length 394;
Best Local Similarity	50.0%	Pred. No. 32;		
Matches	8;	Conservative	2;	Mismatches
			6;	Indels
			0;	Gaps
			0;	

Qy 4 KTPLTTAAAXAPVVXNA 19
||| ||| : :
Db 315 KTPLTTHLTTPALIDA 330

```

RESULT 19
S05661
muscarinic acetylcholine receptor - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 15-Oct-1999
C:Accession: S05661; A36191

```

R;Onai, T.; Fitzgerald, M.G.; Arakawa, S.; Gocayne, J.D.; Urquhart, D.A.; Hall, L.M.; F.
FEBS Lett. 255, 219-225, 1989
A:Title: Cloning, sequence analysis and chromosome localization of a *Drosophila* muscari.
A:Reference number: S05661; MUID:90005981; PMID:2507354
A:Accession: S05661
A:Accession: S05661
A:Molecule type: mRNA
A:Residues: 1-788 <ONA>
A:Cross-references: EMBL:M23412; NID:g157829; PID:g157830
R;Shapiro, R.A.; Wakimoto, B.T.; Subers, E.M.; Nathanson, N.M.
Proc. Natl. Acad. Sci. U.S.A. 86, 9039-9043, 1989
A:Title: Characterization and functional expression in mammalian cells of genomic and c.
A:Reference number: A36191; MUID:90046926; PMID:2510174

Query Match	49.4%	Score 40;	DB 2;	Length 788;
Best Local Similarity	52.9%;	Pred. No. 69;		
Matches	9;	Mismatches	1;	
	Conservative		7;	Indels 0;
				Gaps 0;

Qy	3	EK	PL	TT	AA	XP	VV	XN	A	19
Db	602	ED	GP	TT	TA	AA	AP	LA	SA	618

RESULT 20

T31674

hypothetical protein 1 - fruit fly (*Drosophila simulans*) retrotransposon ninja (fragment)

C:Species: *Drosophila simulans*

C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000

C:Accession: T31674

R:Ogura, K.; Takechi, S.; Nakayama, T.; Yamamoto, M.

Genes Genet. Syst. 71, 1-8, 1996

A:Title: Molecular structure of the transposable element ninja in *Drosophila simulans*.

A:Reference number: Z21057; MUID:96252483; PMID:8925473

A:Accession: T31674

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1360 <OGU>

A:Cross-references: EMBL:D83207; NID:d1068446; PID:d1020552; PIDN:BAA19771.1

C:Genetics:

A:Cross-references: FlyBase:FBgn0015168

A:Mobile element: retrotransposon ninja

Query Match	49.4%	Score 40;	DB 2;	Length 1360;
Best Local Similarity	47.1%	Pred. No. 1.3e+02;		
Matches	8;	Conservative	2;	Mismatches 7;
				Indels 0;
				Gaps 0;

Qy 3 EKTPLTTAAXAPVVXNA 19
| | | | | : | |
Db 101 EATPITTTASQOPASSGA 117

RESULT 21
JQ2147
OHpl protein - maize
C:Species: Zea mays (maize)
C:Date: 28-Aug-1985 #sequence_revision 07-Oct-1994 #text_change 26-May-2000
C:Accession: JQ2147

R;Pysh, L.D.; Aukerman, M.J.; Schmidt, R.J.
Plant Cell 5, 227-236, 1993
A>Title: OHP1: a maize basic domain/leucine zipper protein that interacts with opaque2.
A:Reference number: JQ2147; MUID:93200807; PMID:8453304
A:Accession: JQ2147
A:Molecule type: mRNA
A:Residues: 1-405 <PVS>
A:Cross-references: GB:I00623; NID:g168529; PIDN:AA33488.1; PID:g168530
A:Experimental source: endosperm
C:Comment: This protein has a basic domain and leucine zipper domain and plays a role in C;Genetics:
C:Gene: OHP1
C:Superfamily: seed storage protein opaque-2; fos/jun DNA-binding domain homology
C:Keywords: leucine zipper
F:207-247/Domain: fos/jun DNA-binding domain homology <FUD>
F:233-292/Region: leucine zipper motif

Query Match 48.8%; Score 39.5; DB 2; Length 405;
Best Local Similarity 48.0%; Pred. No. 41;
Matches 12; Conservative 0; Mismatches 6; Indels 7; Gaps 1;

QY 2 EEKTLTTAAXA-----PVVXNA 19
||| ||| ||| ||| |||
Db 79 ERKTPGTAAAAASSSVDPVEYN 103

RESULT 22

S10772
2-hydroxyomuonic semialdehyde dehydrogenase (EC 1.2.1.-) - Pseudomonas putida
C:Species: Pseudomonas putida
C>Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 11-Jun-1999
C:Accession: S10772
R:Nordlund, I.; Shingler, V.
Biochim. Biophys. Acta 1049, 227-230, 1990
A>Title: Nucleotide sequences of the meta-cleavage pathway enzymes 2-hydroxyomuonic semi
A:Reference number: S10772; MUID:90304229; PMID:2194577
A:Accession: S10772
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-486 <NOR>
A:Cross-references: EMBL:X52805; NID:g45677; PIDN:CAA36992.1; PID:g45679
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: oxidoreductase
F:42-308/Domain: aldehyde dehydrogenase homology <ALDD>

Query Match 48.1%; Score 39; DB 2; Length 486;
Best Local Similarity 52.9%; Pred. No. 61;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 EKPLTTLTAAXAPVVXNA 19
||| ||| ||| ||| |||
Db 181 EETPLTTALLGEVMQA 197

RESULT 23

B87400
aldehyde dehydrogenase [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: B87400
R:Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonin, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A>Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: B87400
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-507 <STO>
A:Cross-references: GB:AE005673; NID:gl3422542; PIDN:AAK23198.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC1216

||:|:|:|:|:|:|:

568 EESPASTOASTPVM 582

RESULT 26

D75489

hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000

C:Accession: D75489

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: D75489

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1658 <WHI>

A:Cross-references: GB:AE001925; GB:AE000513; NID:g6458383; PIDN:AAF10268.1; PID:g645839

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0687

A:Map position: 1

Query Match

Best Local Similarity 47.5%; Score 38.5; DB 2; Length 1658;

Matches 10; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 3 EKTPLTT-AAAPVXN 18

Db 1288 ELTPTTVSATAPVTEN 1304

|:|:|:|:|:|:|:

|:|:|:|:|:|:|:

RESULT 27

S68226

growth-blocking peptide precursor - armyworm (Pseudaletia separata)

C:Species: Pseudaletia separata

C:Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 05-Nov-1999

C:Accession: S68226; A39803

R:Hayakawa, Y.; Ohnishi, A.; Yamanaka, A.; Izumi, S.; Tomino, S.

FEBS Lett. 376, 185-189, 1995

A:Title: Molecular cloning and characterization of cDNA for insect biogenic peptide, gro

A:Reference number: S68226; MUID:96105377; PMID:7498538

A:Accession: S68226

A:Molecule type: mRNA

A:Residues: 1-143 <HAY>

A:Cross-references: GB:S80564; NID:g1245957; PIDN:AAB35742.1; PID:g1245958

R:Hayakawa, Y.

J. Biol. Chem. 266, 7982-7984, 1991

A:Title: Structure of a growth-blocking peptide present in parasitized insect hemolymph.

A:Reference number: A39803; MUID:91217017; PMID:2022627

A:Accession: A39803

A:Molecule type: protein

A:Residues: 121-143, 'YQ' <HA2>

C:Superfamily: paralytic peptide I

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-120/Domain: propeptide #status predicted <PRO>

F:121-143/Product: growth-blocking peptide #status predicted <MAT>

F:127-139/Disulfide bonds: #status predicted

Query Match

Best Local Similarity 46.9%; Score 38; DB 2; Length 143;

Matches 10; Conservative 1; Mismatches 7; Indels 2; Gaps 1;

QY 2 EEKTP--LTTAAXAPVXNA 19

Db 83 DEVTPATTTTAAAPVNA 102

|:|:|:|:|:|:|:

|:|:|:|:|:|:|:

RESULT 28

T40089

hypothetical protein C15H7.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 06-Jan-1995

C:Accession: S40750

R:Smith, A.

cytochrome c1, heme protein precursor - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000

C:Accession: T40089

R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.

submitted to the EMBL Data Library, March 1998

A:Reference number: Z21904

A:Accession: T40089

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-307 <LYN>

A:Cross-references: EMBL:AL022299; PIDN:CAA18395.1; GSPDB:GN00067; SPDB:SPBC29A3.18

A:Experimental source: strain 972h-; cosmid C29A3

C:Genetics:

A:Gene: SPDB:SPBC29A3.18

A:Map position: 2

A:Introns: 20/3

C:Superfamily: cytochrome c1 heme protein; cytochrome c1 heme protein homology

C:Keywords: chromoprotein; heme; iron; metalloprotein

F:71-296/Domain: cytochrome c1 heme protein homology <CIH>

F:102,105/Binding site: heme (Cys) (covalent) #status predicted

F:106/Binding site: heme iron (His) (axial ligand) #status predicted

Query Match

Best Local Similarity 46.9%; Score 38; DB 2; Length 307;

Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 EEKPLTTAAXAPVXN 18

Db 237 EDGTPATTAQAQVYN 253

|:|:|:|:|:|:|:

|:|:|:|:|:|:|:

RESULT 29

A35125

phospholipase D (EC 3.1.4.4) precursor - Corynebacterium pseudotuberculosis

C:Species: Corynebacterium pseudotuberculosis

C:Date: 03-Aug-1990 #sequence_revision 03-Aug-1990 #text_change 22-Oct-1999

C:Accession: A35125; S54397

R:Hodgson, A.L.M.; Bird, P.; Nisbet, I.T.

J. Bacteriol. 172, 1256-1261, 1990

A:Title: Cloning, nucleotide sequence, and expression in Escherichia coli of the phosph

A:Reference number: A35125; MUID:90170833; PMID:2407718

A:Accession: A35125

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-307 <HOD>

R:McNamara, P.J.; Bradley, G.A.; Songer, J.G.

Mol. Microbiol. 12, 921-930, 1994

A:Title: Targeted mutagenesis of the phospholipase D gene results in decreased virulence:

A:Reference number: S54397; MUID:95020614; PMID:7934899

A:Accession: S54397

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-307 <MCN>

A:Cross-references: EMBL:U16587; NID:g289915; PIDN:AAA64910.1; PID:g289916

C:Keywords: phosphoric diester hydrolase

Query Match

Best Local Similarity 46.9%; Score 38; DB 2; Length 307;

Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 6 PLTTAAXAPVXN 18

Db 18 PVGNAAAAPVXN 30

|:|:|:|:|:|:|:

|:|:|:|:|:|:|:

RESULT 30

S40750

hypothetical protein C15H7.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 06-Jan-1995

C:Accession: S40750

R:Smith, A.

submitted to the EMBL Data Library, February 1993

A:Reference number: S40750
A:Accession: S40750
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-348 <SMI>
A:Cross-references: EMBL:222173
C:Genetics:
A:Introns: 170/1; 215/3; 297/3; 343/1

Query Match 46.9%; Score 38; DB 2; Length 348;
Best Local Similarity 66.7%; Pred. No. 63;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 TPLTTAAAPV 16
: |||||
DB 229 SPCTTAASAPSV 240

RESULT 31

E87650

peptidase, M20/M25/M40 family [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: E87650
R:NIERMAN, W.C.; FELDBLYUM, T.V.; PAULSEN, I.T.; NELSON, K.E.; EISEN, J.; HEIDELBERG, J.
B.; LAUB, M.T.; DEBOY, R.T.; DODSON, R.J.; DURKIN, A.S.; GWINN, M.L.; HAFT, D.H.; KOLON
N, J.; ERMOLAEVA, M.; WHITE, O.; SALZBERG, S.L.; SHAPIRO, L.; VENTER, J.C.; FRASER, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: E87650

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-474 <STO>

A:Cross-references: GB:AE005673; NID:g13424925; PIDN:AAK25201.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC3239

Query Match 46.9%; Score 38; DB 2; Length 474;
Best Local Similarity 57.1%; Pred. No. 88;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 6 PLTTAAAPV 19
: |||||
DB 386 PLTPAIMAPIQKNA 399

RESULT 32

S40983

hypothetical protein F59A4.11 - Caenorhabditis elegans (fragment)

C:Species: Caenorhabditis elegans

C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Sep-1997

C:Accession: S40983

R:BERKE, M.

submitted to the EMBL Data Library, February 1992

A:Reference number: S40973

A:Accession: S40983

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-532 <BER>

A:Cross-references: EMBL:Z22179; NID:g297956; PID:g297966

C:Genetics:

A:Introns: 82/3; 128/1; 222/1; 280/3; 397/2; 428/3; 491/2

Query Match 46.9%; Score 38; DB 2; Length 532;
Best Local Similarity 66.7%; Pred. No. 1e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 TPLTTAAAPV 16
: |||||
DB 14 SPCTTAASAPSV 25

RESULT 33

F81313

peptidase (M3 family) Cj1099 [imported] - Campylobacter jejuni (strain NCTC 11168)
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: F81313
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell
Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hy
A:Reference number: AB1250; MUID:20150912; PMID:10688204
A:Accession: F81313
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-573 <PAB>

A:Cross-references: GB:AL111168; NID:g5968444; PIDN:CAB73354.1; PID:g59685

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: Cj1099

C:Superfamily: oligopeptidase F

Query Match 46.9%; Score 38; DB 2; Length 573;
Best Local Similarity 61.5%; Pred. No. 1.1e+02;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 EKTPLTTAAAPV 15
: |||||
DB 385 QNTPLTTAAASV 397

RESULT 34

T25603

hypothetical protein C32E8.10 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T25603

R:GATTUNG, S.

submitted to the EMBL Data Library, February 1997

A:Description: The sequence of C. elegans cosmid C32E8.

A:Reference number: Z20056

A:Accession: T25603

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-684 <GAT>

A:Cross-references: EMBL:U88308; PIDN:AA42327.1; GSPDB:GN00019; CESP:C32E8.10

A:Experimental source: strain Bristol N2; clone C32E8

C:Genetics:

A:Gene: CESP:C32E8.10

A:Map position: 1

A:Introns: 24/2; 59/1; 163/3; 255/3; 282/3; 447/1; 487/1; 541/1; 587/2; 636/1

Query Match 46.9%; Score 38; DB 2; Length 684;
Best Local Similarity 44.4%; Pred. No. 1.3e+02;
Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 BEKPLTTAAAPV 19
: |||||
DB 487 EPENPITAPAPQHNA 504

RESULT 35

D71283

probable translation elongation factor G (fusa-2) - syphilis spirochete

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 02-Feb-2001

C:Accession: D71283

R:FRASER, C.M.; NORRIS, S.J.; WEINSTOCK, G.M.; WHITE, O.; SUTTON, G.G.; DODSON, R.; GWI
RSON, J.; KHALAK, H.; RICHARDSON, D.; HOWELL, J.K.; CHIDAMBARAM, M.; UTTERBACK, T.; MC
THEY, L.; WEIDMAN, J.; SMITH, H.O.; VENTER, J.C.
Science 281, 375-388, 1998

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A71250; MUID:98332770; PMID:9665876

A;Accession: D71283
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-695 <COL>
A;Cross-references: GB:AE001248; GB:AE000520; NID:g3323074; PIDN:AAC65735.1; PID:g332307
A;Experimental source: strain Nichols
C;Genetics:
A;Gene: TP0767
C;Superfamily: translation elongation factor G; translation elongation factor Tu homolog
C;Keywords: GTP binding; nucleotide binding; P-loop
F;9-139/Domain: translation elongation factor Tu homology <ETU>
F;15-22/Region: nucleotide-binding motif A (P-loop)
F;136-139/Region: GTP-binding NKXD motif

Query Match 46.9%; Score 38; DB 2; Length 695;
Best Local Similarity 60.0%; Pred. No. 1.3e+02;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAPV 16
|||:|||||
Db 295 EEPVTLSTDAAPV 309
|||:|||||

RESULT 36
F88561
protein F58A4.11 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: F88561
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: F88561
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-747 <STO>
A;Cross-references: GB:chr_III; PIDN:CAA80127.1; PID:g3874287; GSPDB:GN00021; CESP:F58A4
A;Gene: F58A4.11
A;Map position: 3

Query Match 46.9%; Score 38; DB 2; Length 747;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 TPLTTAAAXAPV 16
|||:|||||
Db 229 SPCTTAASAPV 240
|||:|||||

RESULT 37
S48385
hypothetical protein Y11149c - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 19-Apr-2002
C;Accession: S48385
R;Churcher, C.
submitted to the EMBL Data Library, September 1994
A;Reference number: S48310
A;Accession: S48385
A;Molecule type: DNA
A;Residues: 1-1679 <CHU>
A;Cross-references: GB:247047; EMBL:238059; NID:g603997; PID:g763197; GSPDB:GN00009; MIF
C;Genetics:
A;Gene: SGD:MLP2; MIPS:Y11149c
A;Cross-references: SGD:S0001411
A;Map position: 9L

Query Match 46.9%; Score 38; DB 2; Length 1679;
Best Local Similarity 46.7%; Pred. No. 3.6e+02;

Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAPV 16
|||:|||||
Db 346 EBEATMTTSVSPV 360
|||:|||||

RESULT 38
G86194
hypothetical protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: G86194
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.P.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maity, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: G86194
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-200 <STO>
A;Cross-references: GB:AE005172; NID:g8810460; PIDN:AAF80121.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

Query Match 46.3%; Score 37.5; DB 2; Length 200;
Best Local Similarity 50.0%; Pred. No. 42;
Matches 9; Conservative 3; Mismatches 3; Indels 3; Gaps 1;

QY 2 EKTPLTTAAAXAPVXNA 19
|||:|||||
Db 23 DELTLPAT---APVISSA 37
|||:|||||

RESULT 39
F72332
hypothetical protein - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: F72332
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hicke,
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.
C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome se
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: F72332
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-71 <ARN>
A;Cross-references: GB:AE001747; GB:AE000512; NID:g4981304; PIDN:AAD35863.1; PID:g49813
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM0781

Query Match 45.7%; Score 37; DB 2; Length 71;
Best Local Similarity 40.0%; Pred. No. 16;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAPV 16
|||:|||||
Db 18 QRTTPVSTAGSPVM 32
|||:|||||

RESULT 40
C83492

hypothetical protein PA1228 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C83492
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: C83492
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1116 <STO>
A:Cross-references: GB:AE004552; GB:AE004091; NID:g9947150; PIDN:RAG04617.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1228

Query Match 45.7%; Score 37; DB 2; Length 116;
Best Local Similarity 50.0%; Pred. No. 28;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 6 PLTTAAXAPVXNA 19
| | | | |
Db 54 PVTPTAAPVIDKA 67

Search completed: January 28, 2004, 13:09:55
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 13:02:49 ; Search time 11 Seconds
(without alignments)
81.228 Million cell updates/sec

Title: US-09-359-426C-2

Perfect score: 81

Sequence: 1 XEEKPLTTAAAPVVXNA 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	63	77.8	482	1 CATA_PSEAE	O52762 pseudomonas
2	53	65.4	484	1 CATA_PROMI	P42321 proteus mir
3	45	55.6	128	1 BL7 STRAT	P29342 streptomyces
4	43	53.1	483	1 BCA STRVL	P33569 streptomyces
5	41	50.6	482	1 CATA_BORPE	P48062 bordetella
6	40	49.4	722	1 ACMI_DROME	P16395 drosophila
7	39	48.1	143	1 H2B_AGABI	P78567 agarius bi
8	39	48.1	486	1 DMPC_PESP	P19059 pseudomonas
9	39	48.1	527	1 YB64 YEAST	P38314 saccharomyc
10	39	48.1	560	1 VNUC INBLE	P04665 influenza b
11	39	48.1	815	1 GYRB_MYXXA	O33367 myxococcus
12	38	46.9	143	1 GBP_PSESE	O27913 pseudolatelia
13	38	46.9	307	1 PLD_CORPS	P20826 corynebacte
14	38	46.9	396	1 ASSY_STRMU	O8Cw20 streptococ
15	38	46.9	479	1 CATA_PSEPU	O59714 pseudomonas
16	38	46.9	695	1 EFGI_TREPA	O83748 treponema p
17	38	46.9	747	1 YMHA_CAEEL	P34478 caenorhabdi
18	38	46.9	1679	1 YIO9 YEAST	P40457 saccharomyc
19	37	45.7	366	1 FTZI_PYRWO	O52830 pyrococcus
20	37	45.7	396	1 ASSY_STR3	O8E7n1 streptococ
21	37	45.7	396	1 ASSY_STR5	O8E272 streptococ
22	37	45.7	442	1 SDC3 MOUSE	O64519 mus musculu
23	37	45.7	482	1 CATA_VIBFI	O68146 vibrio fisc
24	37	45.7	498	1 LCYB_CAPAN	O43415 capicum an
25	37	45.7	500	1 LCYB_LYCES	O43503 lycopersico
26	37	45.7	500	1 LCYB_TOBAC	O43578 nicotiana t
27	37	45.7	500	1 PSD3 ANOST	O9u528 anopheles s
28	37	45.7	511	1 PSH4 YEAST	P32872 saccharomyc
29	37	45.7	534	1 YAJE_SCHPO	O09911 schizosacch
30	37	45.7	662	1 MUCI_XENLA	O05049 xenopus lae
31	37	45.7	946	1 FTSC_CAMJE	O46089 campylobact
32	37	45.7	1034	1 GCSP_FLAAN	O49850 flaveria an
33	37	45.7	1064	1 ISK5_HUMAN	O9nq38 homo sapien

34	37	45.7	1567	1 ICEN_XANCT	P18127 xanthomonas
35	37	45.7	3390	1 POLG_DEN3	P27915 d genome po
36	36.5	45.1	235	1 CAVI_CAEEL	Q94051 caenorhabdi
37	36	44.4	158	1 HUNB_DROMM	O46248 drosophila
38	36	44.4	214	1 DAMX_SERMA	P45459 serratia ma
39	36	44.4	322	1 ANT_BPPI	P19655 bacterioph
40	36	44.4	340	1 CMFH_SCHPO	O94620 schizosacch
41	36	44.4	501	1 LCYB_ARATH	Q38933 arabidopsis
42	36	44.4	553	1 SPAP_RAT	O88506 rattus norv
43	36	44.4	567	1 ODP2_HAEIN	P45118 haemophilus
44	36	44.4	583	1 AS15_MOUSE	Q8vhs6 mus musculu
45	36	44.4	626	1 GPBA_HUMAN	P07359 homo sapien

ALIGNMENTS

RESULT 1					
ID	CATA_PSEAE	STANDARD;	PRT;	482 AA.	
AC	O52762;				
DT	30-MAY-2000 (Rel. 39, Created)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Catalase [EC 1.11.1.6].				
GN	KATA OR PA4236.				
OS	Pseudomonas aeruginosa.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;				
OC	Pseudomonadaceae; Pseudomonas.				
OX	NCBI_TaxID=287;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=PRD1;				
RX	MEDLINE=99296583; PubMed=10368148;				
RA	Ma J.-F., Ochsner U.A., Klotz M.G., Nanayakkara V.K., Howell M.L.,				
RA	Johnson Z., Posey J.E., Vasili M.L., Monaco J.J., Hassett D.J.;				
RT	"Bacterioferritin A modulates catalase A (Kata) activity and				
RT	resistance to hydrogen peroxide in Pseudomonas aeruginosa.";				
RL	J. Bacteriol. 181:3730-3742(1999).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=ATCC 15692 / PAO1;				
RX	MEDLINE=20437337; PubMed=10984043;				
RA	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,				
RA	Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,				
RA	Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,				
RA	Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,				
RA	Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.;				
RA	Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;				
RT	"Complete genome sequence of Pseudomonas aeruginosa PAO1, an				
RT	opportunistic pathogen.";				
RL	Nature 406:959-964(2000).				
CC	-!- FUNCTION: Decomposes hydrogen peroxide into water and oxygen;				
CC	serves to protect cells from the toxic effects of hydrogen				
CC	peroxide.				
CC	-!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.				
CC	-!- COFACTOR: HEME GROUP.				
CC	-!- ENZYME REGULATION: BY PEROXIDE AND BFR-BOUND IRON.				
CC	-!- SUBUNIT: HETEROMULTIMER. POSSIBLY AN ALPHA(2)BETA-HETEROTRIMER				
CC	WHERE THE ALPHA SUBUNIT IS A 56 kDa PROTEIN AND THE BETA SUBUNIT A				
CC	45 kDa PROTEIN.				
CC	-!- SIMILARITY: Belongs to the catalase family.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL; AF047025; AAC03118.1; -.				
DR	EMBL; AE004841; AAC07624.1; -.				

DR PIR: B83113; B83113.
 DR HSSP: P42321; 2CAE.
 DR InterPro: IPR002226; Catalase.
 DR Pfam: PF00199; catalase; 1.
 DR PRINTS: PR00067; CATALASE.
 DR ProDom: PD000510; Catalase; 1.
 DR PROSITE: PS00437; CATALASE_1; 1.
 DR PROSITE: PS00438; CATALASE_2; 1.
 DR Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
 KW Complete proteome.
 KW ACT SITE 55 55 BY SIMILARITY.
 FT ACT SITE 128 128 BY SIMILARITY.
 FT ACT SITE 338 338 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
 FT METAL
 SQ SEQUENCE 482 AA; 55589 MW; 845EBA647CAB414 CRC64;
 Query Match 77.8%; Score 63; DB 1; Length 482;
 Best Local Similarity 82.4%; Pred. NO. 0.0031;
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 EKKPLTTAAAPVYN 18
 DB 2 EKKRLTTAAGAPVDN 18
 RESULT 2
 CATA_PROMI STANDARD; PRT; 484 AA.
 ID CATA_PROMI
 AC P42321;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Catalase (EC 1.11.1.6).
 GN KATA.
 OS Proteus mirabilis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Proteus.
 OX NCBI_TaxID=584;
 RN [1]
 RP SEQUENCE, AND SEQUENCE OF 1-305 FROM N.A.
 RC STRAIN=PR;
 RX MEDLINE=95305957; PubMed=7786407;
 RA Buzy A., Bracchi V., Stergiades R., Chroboczek J., Thibault P.,
 RA Gagnon J., Jouve H.-M., Hudry-Clergeon G.;
 RT "Complete amino acid sequence of Proteus mirabilis PR catalase.
 RT Occurrence of a methionine sulfone in the close proximity of the
 RT active site.";
 RL J. Protein Chem. 14:59-72(1995).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RC STRAIN=PR;
 RX MEDLINE=95311317; PubMed=7791219;
 RA Guet P., Jouve H.-M., Dideberg O.;
 RT "Crystal structure of Proteus mirabilis PR catalase with and without
 RT bound NADPH.";
 RL J. Mol. Biol. 249:933-954(1995).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS), AND ABSORPTION SPECTROSCOPY.
 RX MEDLINE=97057535; PubMed=8901874;
 RA Guet P., Jouve H.-M., Williams P.A., Andersson I., Andreoletti P.,
 RA Nussbaum L., Hajdu J.;
 RT "Ferryl intermediates of catalase captured by time-resolved
 RT Weissenberg crystallography and UV-VIS spectroscopy.";
 RL Nat. Struct. Biol. 3:951-956(1996).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS), AND EPR SPECTROSCOPY.
 RX MEDLINE=2374619; PubMed=12486720;
 RA Andreoletti P., Sainz G., Jacquinet M., Gagnon J., Jouve H.-M.;
 RT "High-resolution structure and biochemical properties of a recombinant
 RT Proteus mirabilis catalase depleted in iron.";
 RL Proteins 50:261-271(2003).
 CC -!- FUNCTION: Decomposes hydrogen peroxide into water and oxygen;
 CC serves to protect cells from the toxic effects of hydrogen
 CC peroxide.

CC -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
 CC -!- COFACTOR: HEME GROUP AND NADP.
 CC -!- SUBUNIT: Homotetramer.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- MASS SPECTROMETRY: MW=55643; MW_ERR=5; METHOD=Electrospray.
 CC -!- SIMILARITY: Belongs to the catalase family.
 DR PIR: A58663; A58663.
 DR PDB: 1M85; 14-AUG-02.
 DR PDB: 1MQF; 09-OCT-02.
 DR PDB: 2CAG; 07-DEC-96.
 DR PDB: 2CAH; 11-JAN-97.
 DR PDB: 1E93; 13-OCT-00.
 DR InterPro: IPR002226; Catalase.
 DR Pfam: PF00199; catalase; 1.
 DR PRINTS: PR00067; CATALASE.
 DR ProDom: PD000510; Catalase; 1.
 DR PROSITE: PS00437; CATALASE_1; 1.
 DR PROSITE: PS00438; CATALASE_2; 1.
 KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide; NADP;
 KW 3D-structure.
 FT MOD_RES 53 53 METHIONINE SULFONE.
 FT ACT_SITE 54 54
 FT ACT_SITE 127 127
 FT METAL 337 337 IRON (HEME AXIAL LIGAND).
 FT STRAND 7 7
 FT TURN 9 10
 FT STRAND 13 13
 FT STRAND 22 22
 FT TURN 25 26
 FT STRAND 29 29
 FT TURN 31 32
 FT HELIX 34 43
 FT TURN 44 45
 FT TURN 57 66
 FT TURN 71 73
 FT HELIX 77 79
 FT TURN 82 83
 FT TURN 85 93
 FT TURN 99 100
 FT STRAND 110 117
 FT TURN 118 119
 FT STRAND 120 127
 FT TURN 137 138
 FT HELIX 139 147
 FT STRAND 149 149
 FT TURN 151 153
 FT STRAND 156 156
 FT HELIX 158 165
 FT TURN 166 167
 FT HELIX 169 171
 FT HELIX 172 179
 FT HELIX 181 183
 FT STRAND 185 185
 FT TURN 188 189
 FT STRAND 193 194
 FT STRAND 199 202
 FT TURN 204 205
 FT STRAND 208 217
 FT TURN 218 219
 FT STRAND 223 223
 FT HELIX 226 235
 FT TURN 237 238
 FT TURN 239 250
 FT TURN 251 251
 FT STRAND 255 264
 FT HELIX 265 269
 FT TURN 270 270
 FT TURN 275 276
 FT TURN 278 279
 FT TURN 284 286
 FT STRAND 289 299
 FT HELIX 304 307
 FT TURN 308 310

```
FT TURN 315 316
FT TURN 320 321
FT STRAND 322 323
FT HELIX 328 344
FT TURN 345 345
FT TURN 347 348
FT HELIX 349 351
FT HELIX 353 355
FT TURN 369 370
FT TURN 386 387
FT HELIX 394 396
FT HELIX 412 414
FT HELIX 420 426
FT TURN 427 428
FT HELIX 431 445
FT TURN 446 447
FT HELIX 450 463
FT HELIX 465 476
FT TURN 477 477
SQ SEQUENCE 484 AA; 55614 MW; ADC25F3CB41F5C50 CRC64;

Query Match 65.4%; Score 53; DB 1; Length 484;
Best Local Similarity 75.0%; Pred. No. 0.15;
Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 EKTPLTAAAXAPVXN 18
   ||| ||||| |||||
Db 2 EKKKLTTAAGAPVDN 17

RESULT 3
RL7 STRAT STANDARD; PRT; 128 AA.
AC P29342;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L7/L12.
GN RPL7.
OS Streptomyces antibioticus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1890;
RN [1]
RP SEQUENCE FROM N.A., PubMed=1511874;
RX MEDLINE=92380478; PubMed=1511874;
RA Parra F., Blanco G., Alonso J.M., Balbin M., Mendez C., Salas J.A.;
RT "Cloning and sequence of a gene encoding the L7/L12 ribosomal protein
equivalent of Streptomyces antibioticus.";
RL Gene 118:127-129(1992).
CC -!- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS
INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR
ACCURATE TRANSLATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; M89911; AAA26811.1; -
CC PIR; JCI273; JCI273.
CC HSSP; P02392; 1CTF.
CC HAMAP; MF 00368; -; 1.
CC InterPro; IPR000206; Ribosomal L12.
CC Pfam; PF00542; Ribosomal L12; 1.
CC ProDom; PD001326; Ribosomal L12; 1.
CC TIGRFAMs; TIGR00855; L12; 1.
CC KWRibosomal protein.
SQ SEQUENCE 128 AA; 13272 MW; F5C3EE4F45D606E8 CRC64;

Query Match 55.6%; Score 45; DB 1; Length 128;
Best Local Similarity 66.7%; Pred. No. 0.89;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 EKTPLTAAAXAPV 16
   ||| :| |||
Db 28 EEKFDVTAATAAPV 42

RESULT 4
BCA_STRLV STANDARD; PRT; 483 AA.
AC P33569;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bromoperoxidase-catalase (EC 1.11.1.1.-).
GN BCA.
OS Streptomyces violaceus (Streptomyces venezuelae).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1936;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 10712 / ISP5230;
RX MEDLINE=97022081; PubMed=8868441;
RA Facey S., Gross F., Vining L.C., Yang K., van Pee K.-H.;
RT "Cloning, sequencing and disruption of a bromoperoxidase-catalase
gene in Streptomyces venezuelae: evidence that it is not required for
chlorination in chloramphenicol biosynthesis.";
RL Microbiology 142:657-665(1996).
CC -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC -!- SIMILARITY: Belongs to the catalase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; X74791; CAA52796.1; -
CC PIR; S37055; S37055.
CC HSSP; P42321; 2CAE.
CC InterPro; IPR002226; Catalase.
CC Pfam; PF00199; catalase; 1.
CC PRINTS; PR00067; CATALASE.
CC ProDom; PD000510; Catalase; 1.
CC PROSITE; PS00437; CATALASE_1; 1.
CC PROSITE; PS00438; CATALASE_2; 1.
CC Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
FT ACT_SITE 54 54 BY SIMILARITY.
FT ACT_SITE 127 127 BY SIMILARITY.
FT METAL 337 337 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 483 AA; 54087 MW; B58CF8230B8A2F55 CRC64;

Query Match 53.1%; Score 43; DB 1; Length 483;
Best Local Similarity 69.2%; Pred. No. 7.5;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 6 PLTTAAXAPVXN 18
   ||| |||||
Db 5 PLTTAGAPVDN 17

RESULT 5
CATA_BORPE STANDARD; PRT; 482 AA.
ID CATA_BORPE
AC P48062;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
SQ
```

```

DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Catalase (EC 1.11.1.6).
GN KATA.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP504;
RA Deshazer D., Wood G.E., Friedman R.L.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Decomposes hydrogen peroxide into water and oxygen;
CC serves to protect cells from the toxic effects of hydrogen
CC peroxide.
CC -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC -!- COFACTOR: HEME GROUP.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SIMILARITY: Belongs to the catalase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U07800; AAA18481.1; -.
DR PIR; S60757; S60757.
DR HSP; P42321; 2CAE.
DR InterPro: IPR002226; Catalase.
DR Pfam; PF00199; catalase; 1.
DR PRINTS; PR00067; CATALASE.
DR ProDom; PD000510; Catalase; 1.
DR PROSITE; PS00437; CATALASE 1; 1.
DR PROSITE; PS00438; CATALASE 2; 1.
KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
FT ACT_SITE 57 57 BY SIMILARITY.
FT ACT_SITE 130 130 BY SIMILARITY.
FT METAL 340 340 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 482 AA; 54508 MW; 7CB73E08975C219F CRC64;

Query Match 50.6%; Score 41; DB 1; Length 482;
Best Local Similarity 75.08; Pred. No. 16;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 LTTAAAPVVKX 18
   |||||
DB 9 LTTAAGPVDN 20

RESULT 6
ACM1 DROME
ID ACM1 DROME STANDARD; PRT; 722 AA.
AC P16335;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Muscarinic acetylcholine receptor DM1.
GN ACRC OR MACR-60C.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RA MEDLINE=90046926; PubMed=2510174;
RA Shapiro R.A., Wakimoto B.T., Subers E.M., Nathanson N.M.;
RT "Characterization and functional expression in mammalian cells of
RT genomic and cDNA clones encoding a Drosophila muscarinic

```

```

RT acetylcholine receptor.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9039-9043 (1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9005981; PubMed=2507354;
RA Onai T., Fitzgerald M.G., Arakawa S., Gocayne J.D., Urquhart D.A.,
RA Hall L.M., Fraser C.M., McCombie W.R., Venter J.C.;
RT "Cloning, sequence analysis and chromosome localization of a
RT Drosophila muscarinic acetylcholine receptor.";
RL FEBS Lett. 255:219-225 (1989).
CC -!- FUNCTION: THE MUSCARINIC ACETYLCHOLINE RECEPTOR MEDIATES VARIOUS
CC CELLULAR RESPONSES, INCLUDING INHIBITION OF ADENYLATE CYCLASE,
CC BREAKDOWN OF PHOSPHOINOSITIDES & MODULATION OF POTASSIUM CHANNELS
CC THROUGH THE ACTION OF G PROTEINS. PRIMARY TRANSDUCING EFFECT IS
CC PI TURNOVER.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M27495; AAA85449.1; -.
DR EMBL; M23412; AAA28676.1; ALT_INIT.
DR PIR; S05661; S05661.
DR Flybase; FBgn0000037; mAcR-60C.
DR GO; GO:0005886; C:plasma membrane; NAS.
DR GO; GO:0004981; F:muscarinic acetylcholine receptor activity; IDA.
DR InterPro: IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRRHODOPSN.
DR PROSITE; PS00237; G PROTEIN RECP Fl_1; 1.
DR PROSITE; PS0262; G PROTEIN RECP Fl_2; 1.
KW Postsynaptic membrane; Ionic channel; Glycoprotein; Transmembrane;
KW Phosphorylation; Multigene family; G-protein coupled receptor.
FT DOMAIN 1 26 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 27 49 1 (POTENTIAL). (POTENTIAL).
FT DOMAIN 50 60 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 61 81 2 (POTENTIAL).
FT DOMAIN 82 100 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 101 120 3 (POTENTIAL).
FT DOMAIN 121 140 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 141 162 4 (POTENTIAL).
FT DOMAIN 163 184 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 185 208 5 (POTENTIAL).
FT DOMAIN 209 234 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 235 250 6 (POTENTIAL).
FT DOMAIN 251 266 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 267 272 7 (POTENTIAL).
FT DOMAIN 273 288 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 289 304 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 310 T -> A (IN REF. 2).
FT CONFLICT 311 319 N -> S (IN REF. 2).
FT CONFLICT 320 336 G -> R (IN REF. 2).
FT CONFLICT 337 347 A -> AA (IN REF. 2).
FT CONFLICT 348 350 P -> G (IN REF. 2).
FT CONFLICT 351 362 MISSING (IN REF. 2).
FT CONFLICT 363 381 G -> A (IN REF. 2).
FT CONFLICT 382 381 T -> A (IN REF. 2).
FT CONFLICT 382 381 T -> A (IN REF. 2).
FT CONFLICT 382 381 P -> VG (IN REF. 2).
FT CONFLICT 382 381 P -> AR (IN REF. 2).
FT CONFLICT 382 381 VLI -> CXS (IN REF. 1).
FT CONFLICT 382 381 S -> C (IN REF. 2).
FT CONFLICT 382 381 EGMVRGVYN -> DFVAASTIR (IN REF. 1).
FT CONFLICT 382 381 S -> C (IN REF. 2).
FT CONFLICT 382 381 S -> C (IN REF. 2).
SQ SEQUENCE 722 AA; 78237 MW; FB35417CB220A61 CRC64;

Query Match 49.4%; Score 40; DB 1; Length 722;
Best Local Similarity 52.9%; Pred. No. 36;

```

Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 EKTPLTTAAAXAPVXNA 19

DB 538 EDGPTTTAAAPLASAA 554

RESULT 7

ID H2B AGABI STANDARD; PRT; 143 AA.

AC P78567;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Histone H2B.

GN HTBA.

OS Agaricus bisporus (Common mushroom).

OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;

OC Agaricales; Agaricaceae; Agaricus.

OX NCBI_TaxID=5341;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Horsst UL;

RX MEDLINE=97111990; PubMed=8953726;

RA Sonnenberg A.S.M., de Groot P.W.J., Schaap P.J., Baars J.J.P.,

RA Visser J., van Griensven L.J.L.D.;

RT "Isolation of expressed sequence tags of Agaricus bisporus and their

RT assignment to chromosomes.;"

RL Appl Environ. Microbiol. 62:4542-4547(1996).

CC -1- SUBUNIT: The nucleosome is an octamer containing two molecules

CC each of H2A, H2B, H3 and H4. The octamer wraps approximately 146

CC BP of DNA.

CC -1- SIMILARITY: Belongs to the histone H2B family.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

CC EMBL; X94188; CAA63898.1; -

CC InterPro; IPR004822; Histone core.

CC InterPro; IPR000558; Histone_H2B.

CC Pfam; PF00125; histone; 1.

CC PRINTS; PR00621; HISTONEH2B.

CC ProDom; PD000497; Histone_H2B; 1.

CC SMART; SM00427; H2B; 1.

CC PROSITE; PS00357; HISTONE_H2B; 1.

CC Chromosomal protein; Nucleosome core; Nuclear protein; DNA-binding.

CC SEQUENCE 143 AA; 15165 MW; 69640102E31F4B56 CRC64;

Query Match 48.1%; Score 39; DB 1; Length 143;

Best Local Similarity 50.0%; Pred. No. 10;

Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 KTLPLTTAAAXAPVXNA 19

DB 11 KAPASTASKAPVKSDA 26

RESULT 8

ID DMPC_PSESP STANDARD; PRT; 486 AA.

AC P19059;

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 2-hydroxyxyuconic semialdehyde dehydrogenase (EC 1.2.1.1-) (HMSD).

GN DMPC.

OS Pseudomonas sp. (strain CF600).

OG Plasmid pVII50.
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90304229; PubMed=2194577;
RA Nordlund I., Shingler V.;
RT "Nucleotide sequences of the meta-cleavage pathway enzymes 2-
RT hydroxyxyuconic semialdehyde dehydrogenase and 2-hydroxyxyuconic
RT semialdehyde hydrolase from Pseudomonas CF600.;"
RL Biochim. Biophys. Acta 1049:227-230(1990).
CC -1- FUNCTION: 2-HYDROXYXYUCONIC ACID SEMIALDEHYDE CAN BE CONVERTED TO
CC 2-HYDROXYXYUCONIC ACID SEMIALDEHYDE EITHER DIRECTLY BY THE ACTION OF
CC 2-HYDROXYXYUCONIC SEMIALDEHYDE HYDROLASE (HMSH) OR BY THE ACTION OF
CC THREE SEQUENTIAL ENZYMES, THE FIRST OF WHICH IS HMSD.
CC -1- PATHWAY: META-CLEAVAGE PATHWAY FOR THE DEGRADATION OF PHENOLS,
CC CRESOLS AND CATECHOL. PHENOL METABOLISM.
CC -1- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

CC EMBL; X52805; CAA36992.1; -

CC HSPSP; P20000; LAG8.

CC InterPro; IPR002086; Aldehyde_dehydr.

CC Pfam; PF00171; aldehyd; 1.

CC PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.

CC Aromatic hydrocarbons catabolism; Oxidoreductase; NAD; Plasmid.

FT ACT_SITE 254 254 BY SIMILARITY.

FT ACT_SITE 286 286 BY SIMILARITY.

SQ SEQUENCE 486 AA; 51682 MW; F906FCA64185AA68 CRC64;

Query Match 48.1%; Score 39; DB 1; Length 486;

Best Local Similarity 52.9%; Pred. No. 36;

Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 EKTPLTTAAAXAPVXNA 19

DB 181 EETPLTTALLGEVMOAA 197

RESULT 9

ID YB64_YEAST STANDARD; PRT; 527 AA.

AC P38314;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Hypothetical 57.2 kDa protein in MET8-HPC2 intergenic region.

GN YBR214W OR YBR1501.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=S288c;

RA Dubois E., el Bakkoury M., Glandsdorff N., Messenguy F., Pierard A.,

RA Scherens B., Vierendeels F.;

RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: TO YEAST YGL056C AND S.POMBE SDS23.

CC -1- SIMILARITY: Contains 2 CBS domains.

```
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; Z36083; CAA85178.1; -.
DR PIR; S46088; S46088.
DR SGD; S000418; SDS24.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0006259; P:DNA metabolism; IGI.
DR GO; GO:0006897; P:endocytosis; IMP.
DR GO; GO:0007126; P:meliosis; IEP.
DR GO; GO:0007126; P:meliosis; IEP.
DR InterPro; IPR000644; CBS_domain.
DR Pfam; PF00571; CBS; 4.
DR SMART; SM00116; CBS; 2.
KW Hydrothermal protein; Repeat; CBS domain.
FT DOMAIN 196 251 CBS 1.
FT DOMAIN 283 335 CBS 2.
SQ SEQUENCE 527 AA; 57187 MW; DC2741550A69C154 CRC64;

Query Match 48.1%; Score 39; DB 1; Length 527;
Best Local Similarity 61.5%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAP 14
   |||||
Db 52 EESTPTATAAAP 64

RESULT 10
VNUC_INBLE STANDARD; PRT; 560 AA.
AC P04665;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE Nucleoprotein.
GN NP.
OS Influenza B virus (strain B/Lee/40).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenzavirus B.
OC NCBI_TaxID=11535;
RX MEDLINE=84174071; PubMed=6324462;
RA Briedis D.J., Tobin M.;
RT "Influenza B virus genome: complete nucleotide sequence of the
RT influenza B/lee/40 virus genome RNA segment 5 encoding the
RT nucleoprotein and comparison with the B/Singapore/222/79
RT nucleoprotein."
RL Virology 133:448-455(1984).
CC -!- FUNCTION: ENCAPSIDATES THE NEGATIVE STRAND VIRAL RNA.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; K01395; AAA43689.1; -.
DR InterPro; IPR002141; Flu_NP.
DR Pfam; PF00506; Flu_NP; 1.
KW Nucleoprotein.
SQ SEQUENCE 560 AA; 61770 MW; 74C4FEAF9E75A695 CRC64;

Query Match 48.1%; Score 39; DB 1; Length 560;
Best Local Similarity 44.4%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAPVVKNA 19
   |||||
Db 19 EELTGTSGATRPILKPA 36

RESULT 11
GYRB_MYXA STANDARD; PRT; 815 AA.
AC O33367;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA gyrase subunit B (EC 5.99.1.3).
DE GYRB.
GN Myxococcus xanthus.
OS Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Cystobacterineae; Myxococcaceae; Myxococcus.
OC NCBI_TaxID=34;
RN [1]_TaxID=34;
RP SEQUENCE FROM N.A.
RC STRAIN=ER-15;
RX MEDLINE=98304088; PubMed=9639935;
RA Paitan Y., Boulton N., Ron E.Z., Rosenberg E., Orr E.;
RT "Molecular analysis of the DNA gyrase gene from Myxococcus xanthus.";
RL Microbiology 144:1641-1647(1998).
CC -!- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-
CC STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE
CC INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED
CC DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.
CC -!- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC -!- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA
CC BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE
CC ENZYME FORMS AN A2B2 TETRAMER.
CC -!- SIMILARITY: Belongs to the type II topoisomerase family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AJ000543; CAA04176.1; -.
DR HSSP; P06982; 1AJ6.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR002288; DNA_gyraseB_C.
DR InterPro; IPR000565; DNA_gyraseB.
DR InterPro; IPR001241; DNA_topoisomII.
DR InterPro; IPR006171; Toprim_dom.
DR Pfam; PF00204; DNA_gyraseB; 1.
DR Pfam; PF00986; DNA_gyraseB_C; 1.
DR Pfam; PF02518; DNA_gyraseB_C; 1.
DR Pfam; PF01751; Toprim_1.
DR PRINTS; PR00418; TP12FAMILY.
DR ProDom; PD149633; DNA_gyraseB_C; 1.
DR SMART; SM00387; HATPase_C; 1.
DR SMART; SM00433; TOP2C; 1.
DR TIGRFAMs; TIGR01059; gyraseB; 1.
DR PROSITE; PS00177; TOPOISOMERASE_II; 1.
KW Topoisomerase; Isomerase; ATP-binding.
SQ SEQUENCE 815 AA; 89636 MW; 3862685FBB805B32 CRC64;

Query Match 48.1%; Score 39; DB 1; Length 815;
Best Local Similarity 66.7%; Pred. No. 61;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 EKTPLTTAAAXAP 14
   |||||
Db 2 EKTPTGSAVAP 13
```



```

RESULT 12
GBP_PSESE
ID GBP_PSESE STANDARD; PRT; 143 AA.
AC Q27913;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Growth-blocking peptide precursor (GBP).
OS Pseudaletia separata (Armyworm).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;
OC Noctuidae; Hadeninae; Pseudaletia.
OX NCBI_TaxID=7105;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96105377; PubMed=7498538;
RA Hayakawa Y., Ohnishi A., Yamanaka A., Izumi S., Tomino S.;
RT "Molecular cloning and characterization of cDNA for insect biogenic
RT peptide, growth-blocking peptide.";
RL FEBS Lett. 376:185-189(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98316655; PubMed=9654083;
RA Hayakawa Y., Noguchi H.;
RT "Growth-blocking peptide expressed in the insect nervous system:
RT cloning and functional characterization.";
RL Eur. J. Biochem. 253:810-816(1998).
RN [3]
RP STRUCTURE BY NMR OF 121-143.
RC "TISSUE=Hemolymph;
RX MEDLINE=99107831; PubMed=9890941;
RA Aizawa T., Fujitani N., Hayakawa Y., Ohnishi A., Ohkubo T., Kumaki Y.,
RA Kawano K., Hikichi K., Nitta K.;
RT "Solution structure of an insect growth factor, growth-blocking
RT peptide.";
RL J. Biol. Chem. 274:1887-1890(1999).
CC -1- FUNCTION: BIOGENIC PEPTIDE THAT PREVENTS, IN LEPIDOPTERAN, THE
CC ONSET OF METAMORPHOSIS FROM LARVA TO PUPA. THIS GROWTH-BLOCKING
CC PEPTIDE HAS REPRESSIVE ACTIVITY AGAINST JUVENILE HORMONE ESTERASE.
CC -1- SIMILARITY: BELONGS TO THE GBP / PSP1 / PARALYTIC PEPTIDE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S80564; AAB35742.1; -.
CC EMBL; AB012294; BAA32793.1; -.
CC PIR; S68226; S68226.
CC PDB; 1BOF; 09-DEC-98.
CC InterPro; IPR003463; GBP_PSP.
CC Pfam; PF02425; GBP_PSP; 1.
CC ProDom; PD06507; GBP_PSP; 1.
KW Hemolymph; Signal; 3D-structure.
FT SIGNAL 1 22 POTENTIAL.
FT PROPEP 23 120 GROWTH-BLOCKING PEPTIDE.
FT PEPTIDE 121 143
FT DISULFID 127 139
FT TURN 129 130
FT STRAND 131 133
FT STRAND 139 141
SQ SEQUENCE 143 AA; 15256 MW; DEBFB527956840EB CRC64;

Query Match 46.9%; Score 38; DB 1; Length 143;
Best Local Similarity 50.0%; Pred. No. 15;
Matches 10; Conservative 1; Mismatches 7; Indels 2; Gaps 1;
QY 2 EEXTP--LTTAAAPVAVVNA 19

RESULT 13
PLD_CORPS
ID PLD_CORPS STANDARD; PRT; 307 AA.
AC P20626; Q59314;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Phospholipase D precursor (EC 3.1.4.4) (PLD) (Choline phosphatase).
GN PLD.
OS Corynebacterium pseudotuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1719;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-47.
RX MEDLINE=90170833; PubMed=2407718;
RA Hodgson A.L.M., Bird P., Nisbet I.T.;
RT "Cloning, nucleotide sequence, and expression in Escherichia coli of
RT the phospholipase D gene from Corynebacterium pseudotuberculosis.";
RL J. Bacteriol. 172:1256-1261(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Biovar Ovis / Isolate Whetten 1;
RX MEDLINE=90093451; PubMed=2403529;
RA Songer J.G., Libby S.J., Iandolo J.J., Cuevas W.A.;
RT "Cloning and expression of the phospholipase D gene from
RT Corynebacterium pseudotuberculosis in Escherichia coli.";
RL Infect. Immun. 58:131-136(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=Biovar equi / Isolate 155;
RX MEDLINE=95255653; PubMed=7737503;
RA McNamara P.J., Cuevas W.A., Songer J.G.;
RT "Toxic phospholipases D of Corynebacterium pseudotuberculosis, C.
RT ulcerans and Arcanobacterium haemolyticum: cloning and sequence
RT homology.";
RL Gene 156:113-118(1995).
RN [4]
RP MUTAGENESIS.
RX STRAIN=Biovar Ovis / Isolate Whetten 1;
RX MEDLINE=95020614; PubMed=7934899;
RA McNamara P.J., Bradley G.A., Songer J.G.;
RT "Targeted mutagenesis of the phospholipase D gene results in
RT decreased virulence of Corynebacterium pseudotuberculosis.";
RL Mol. Microbiol. 12:921-930(1994).
CC -1- FUNCTION: VIRULENCE FACTOR AFFECTING BACTERIAL DISSEMINATION AND
CC SURVIVAL WITHIN THE HOST. HAS MAGNESIUM-DEPENDENT SPHINGOMYELINASE
CC AND HEMOLYTIC ACTIVITY.
CC -1- CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = choline + a
CC phosphatidate.
CC -1- SIMILARITY: TO OTHER CORYNEBACTERIUM PHOSPHOLIPASES D.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L16587; AAA64910.1; -.
CC EMBL; L16586; AAA99867.1; -.
CC PIR; A35125; A35125.
KW Hydrolase; Lipid degradation; Magnesium; Virulence; Signal.
FT SIGNAL 1 24
FT CHAIN 25 307 PHOSPHOLIPASE D.
FT ACT_SITE 44 44 POTENTIAL.
FT VARIANT 5 6 VV -> FA (IN BIOVAR EQUI / ISOLATE 155).
FT VARIANT 8 8 F -> L (IN BIOVAR EQUI / ISOLATE 155).

```

```
FT VARIANT      189 189 E -> G (IN BIOVAR EQUI / ISOLATE 155).
FT VARIANT      205 205 N -> D (IN BIOVAR EQUI / ISOLATE 155).
FT VARIANT      270 270 I -> M (IN BIOVAR EQUI / ISOLATE 155).
FT VARIANT      277 277 A -> P (IN BIOVAR EQUI / ISOLATE 155).
SQ SEQUENCE     307 AA; 33884 MW; D3B1334E6FC99875 CRC64;

Query Match      46.9%; Score 38; DB 1; Length 307;
Best Local Similarity 61.5%; Pred. No. 33;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 6 PLTTAAXAPVWVN 18
   |: |||||
Db 18 PVGNAAAAPVWVN 30

RESULT 14
ASSY_STRMU
ID ASSY_STRMU STANDARD; PRT; 396 AA.
AC Q8CW20;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Argininosuccinate synthase (EC 6.3.4.5) (Citruiline--aspartate
DE ligase).
GN ARGG OR SMU.334.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=UAI59 / ATCC 700610 / Serotype C;
RX MEDLINE=22395063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
RT pathogen."
RL Proc.Natl.Acad.Sci.U.S.A. 99:14434-14439(2002).
CC -|- CATALYTIC ACTIVITY: ATP + L-citrulline + L-aspartate = AMP +
CC diphosphate + L-argininosuccinate.
CC -|- PATHWAY: Arginine biosynthesis; seventh step.
CC -|- SUBUNIT: Homotetramer (By similarity).
CC -|- SIMILARITY: Belongs to the argininosuccinate synthase family.
CC Subfamily 1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE014881; AAN58093.1; -.
CC DR HAMAP; MF 00005; -.
CC DR Pfam; PF00764; Arginosuc_synth; 1.
CC DR ProDom; PD003544; Arginosuc_synth; 1.
CC DR TIGRFAMs; TIGR00032; argG; 1.
CC DR PROSITE; PS00564; ARGININOSUCCIN SYN 1; 1.
CC DR PROSITE; PS00565; ARGININOSUCCIN SYN 2; 1.
CC DR Arginine biosynthesis; Ligase; ATP-binding; Complete proteome.
CC SEQUENCE 396 AA; 43913 MW; 44A7C6D56865137E CRC64;

Query Match      46.9%; Score 38; DB 1; Length 396;
Best Local Similarity 46.7%; Pred. No. 43;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 EEKTLTTAAXAPVW 16
   |: |||||
Db 82 EQKPLVLSALSRPVI 96
```

```
RESULT 15
CATAL_PSEPU
ID CATA_PSEPU STANDARD; PRT; 479 AA.
AC Q59714;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Catalase (EC 1.11.1.6).
GN KATA OR CATA.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Corvallis;
RX MEDLINE=98019091; PubMed=9358059;
RA Kim Y.C., Miller C.D., Anderson A.J.;
RT "Identification of adjacent genes encoding the major catalase and a
RT bacterioferritin from the plant-beneficial bacterium Pseudomonas
RT putida."
RL Gene 199:219-224(1997).
CC -|- FUNCTION: Decomposes hydrogen peroxide into water and oxygen;
CC serves to protect cells from the toxic effects of hydrogen
CC peroxide.
CC -|- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC -|- COFACTOR: HEME GROUP.
CC -|- ENZYME REGULATION: ACTIVATED BY PEROXIDE.
CC -|- SIMILARITY: Belongs to the catalase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U63511; AAB88219.1; -.
CC DR HSSP; P42321; 2CAE.
CC DR InterPro; IPR002226; Catalase.
CC DR Pfam; PF00199; catalase; 1.
CC DR PRINTS; PR00067; CATALASE.
CC DR ProDom; PD000510; Catalase; 1.
CC DR PROSITE; PS00437; CATALASE 1; 1.
CC DR PROSITE; PS00438; CATALASE 2; 1.
CC DR Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
CC ACT SITE 53 53 BY SIMILARITY.
CC FT ACT SITE 126 126 BY SIMILARITY.
CC FT METAL 336 336 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
CC SEQUENCE 479 AA; 53381 MW; EPE3CBDE67778571 CRC64;

Query Match      46.9%; Score 38; DB 1; Length 479;
Best Local Similarity 66.7%; Pred. No. 52;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 7 LTTAAXAPVWVN 18
   |||||
Db 5 LTTASGAPVADN 16

RESULT 16
EFGI_TREPA
ID EFGI_TREPA STANDARD; PRT; 695 AA.
AC O83748;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Elongation factor G 1 (EF-G 1).
GN FUSA OR FUSA-2 OR TP0767.
OS Treponema pallidum.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
```

```

RN SEQUENCE FROM N.A.
RP STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RA "Complete genome sequence of Treponema pallidum, the syphilis
RT Spirochete.";
RL Science 281:375-388 (1998).
CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLOCATION
CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
CC RIBOSOME.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-G/EF-2 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF001248; AAC65735.1; -
CC PIR; D71283; D71283.
CC HSSP; P13551; 2EFG.
CC TIGR; TP0767; -
CC HAMAP; MF 00054; -; 1.
CC InterPro; IPR004540; EF-G.
CC InterPro; IPR000795; EF_GTPbind.
CC InterPro; IPR000640; EFG_C.
CC InterPro; IPR005517; EFG_IV.
CC InterPro; IPR004161; EFTU_D2.
CC InterPro; IPR005225; Small_GTP.
CC Pfam; PF00679; EFG_C; 1.
CC Pfam; PF03764; EFG_IV; 1.
CC Pfam; PF00009; GTP_EFTU; 1.
CC Pfam; PF03144; GTP_EFTU_D2; 1.
CC PRINTS; PR00315; ELONGATNCT.
CC TIGRFAMS; TIGR00484; EF-G; 1.
CC TIGRFAMS; TIGR00231; small_GTP; 1.
CC PROSITE; PS00301; EFATOR_GTP; 1.
CC Elongation factor; Protein biosynthesis; GTP-binding;
KW Multigene family; Complete proteome.
FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 82 86 GTP (BY SIMILARITY).
FT NP_BIND 136 139 GTP (BY SIMILARITY).
SQ SEQUENCE 695 AA; 76831 MW; 46529989BFE97E4E CRC64;

Query Match 46.9%; Score 38; DB 1; Length 695;
Best Local Similarity 60.0%; Pred. No. 76;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 EKKPTLTTAAXAPVV 16
DB 295 EEPVTLSTDADAPVV 309

RESULT 17
YMH4_CAEEL
ID YMH4_CAEEL STANDARD; PRT; 747 AA.
AC P34478;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein F58A4.11 in chromosome III.
GN F58A4.11.

```

```

OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fullston L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38 (1994).
CC -!- SIMILARITY: Contains 2 BED-type zinc fingers.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z22173; CAA80127.1; -
CC EMBL; Z22179; CAA80127.1; JOINED.
CC EMBL; Z22179; CAA80171.1; -
CC EMBL; Z22173; CAA80171.1; JOINED.
CC PIR; F88561; F88561.
CC WormPep; F58A4.11; CE00217.
CC InterPro; IPR003656; BED_finger.
CC PROSITE; PS50808; ZF_BED; 2.
KW Hypothetical protein; Repeat; Zinc-finger.
FT ZN_FING 250 298 BED-TYPE 1.
FT ZN_FING 480 529 BED-TYPE 2.
SQ SEQUENCE 747 AA; 83221 MW; 23890C5F8DCFFB15 CRC64;

Query Match 46.9%; Score 38; DB 1; Length 747;
Best Local Similarity 66.7%; Pred. No. 82;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 TPLTTAAXAPVV 16
DB 229 SPCITTAASAPSV 240

RESULT 18
YI09_YEAST
ID YI09_YEAST STANDARD; PRT; 1679 AA.
AC P40457;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical 195.1 kDa protein in DNA43-UBI1 intergenic region.
GN YIL149C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=S288c / AB972;
RX PubMed=9169870;
RA Churcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,
RA Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,

```

DR Pfam: PF03953; tubulin C; 1.
 DR PRINTS; PR00423; CELLDIVISFYZ.
 DR TIGRFAMs; TIGR00065; fte2; 1.
 DR PROSITE; PS01134; FTSZ 1; 1.
 DR PROSITE; PS01135; FTSZ 2; 1.
 DR Cell division; Septation; GTP-binding; Multigene family.
 KW NP_BIND 128 136 GTP (POTENTIAL)
 FT SEQUENCE 366 AA; 39511 MW; 8B40594318F95BFF CRC64;

 Query Match 45.7%; Score 37; DB 1; Length 366;
 Best Local Similarity 46.7%; Pred. No. 59;
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

 QY 2 EKTPLTTAAAPV 16
 ||:||||:|
 DB 345 EETPLETPESP 359

 RESULT 20
 ASSY STRA3 STANDARD; PRT; 396 AA.
 ID AC Q8E7N1;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Argininosuccinate synthase (EC 6.3.4.5) (Citrulline-aspartate
 ligase).
 GN ARGG OR GBS0123.
 OS Streptococcus agalactiae (serotype III).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OC NCBI_TaxID=216495;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=NM316 / Serotype III;
 RX MEDLINE=222422508; PubMed=12354221;
 RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,
 RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
 RA Kunst F.;
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
 invasive neonatal disease";
 RL Mol. Microbiol. 45:1499-1513(2002).
 CC -!- CATALYTIC ACTIVITY: ATP + L-citrulline + L-aspartate = AMP +
 diphosphate + L-argininosuccinate.
 CC -!- PATHWAY: Arginine biosynthesis; seventh step.
 CC -!- SUBUNIT: Homotetramer (By similarity).
 CC -!- SIMILARITY: Belongs to the argininosuccinate synthase family.
 CC Subfamily 1.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL766843; CAD45768.1; -.
 DR Sagalust; gbs0123; -.
 DR HAMAP; MF_00005; -; 1.
 DR InterPro; IPR001518; Arginosuc_synth.
 DR Pfam; PF00764; Arginosuc synth; 1.
 DR ProDom; PD003544; Arginosuc synth; 1.
 DR TIGRFAMs; TIGR00032; argG; 1
 DR PROSITE; PS00564; ARGININOSUCCIN SYN 1; 1.
 DR PROSITE; PS00565; ARGININOSUCCIN SYN 2; 1.
 KW Arginine biosynthesis; ligase; ATP-binding; Complete proteome.
 SQ SEQUENCE 396 AA; 43725 MW; 31DB024B0785AC52 CRC64;

 Query Match 45.7%; Score 37; DB 1; Length 396;
 Best Local Similarity 40.0%; Pred. No. 63;
 Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

```
QY 2 EKTPLTTAAAPV 16
Db 82 EQKPLVLSALSRPII 96

RESULT 21
ASSY_STRAS
ID ASSY_STRAS STANDARD; PRT; 396 AA.
AC Q8E272;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Argininosuccinate synthase (EC 6.3.4.5) (Citrulline-aspartate
ligase).
GN ARGG OR SAGO125.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547;
RA Tettelin H., Masignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
Wessels M.R., Paulsen I.T., Neilson K.E., Margarit I., Read T.D.,
Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
Carly H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
Fraser C.M.;
RA "Complete genome sequence and comparative genomic analysis of an
emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
CC -!- CATALYTIC ACTIVITY: ATP + L-citrulline + L-aspartate = AMP +
diphosphate + L-argininosuccinate.
CC -!- PATHWAY: Arginine biosynthesis; seventh step.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SIMILARITY: Belongs to the argininosuccinate synthase family.
CC Subfamily 1.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AB014197; AAM99033.1; -
DR TIGR; SAGO125; -; 1.
DR HAMAP; MF_00005; -; 1.
DR InterPro; IPR001518; Arginosuc_synth.
DR Pfam; PF00764; Arginosuc_synth; 1.
DR ProDom; PD003544; Arginosuc_synth; 1.
DR TIGRFAMs; TIGR000032; argG; 1.
DR PROSITE; PS00564; ARGININOSUCCIN SYN 1; 1.
DR PROSITE; PS00565; ARGININOSUCCIN SYN 2; 1.
KW Arginine biosynthesis; Ligase; ATP-binding; Complete proteome.
SQ SEQUENCE 396 AA; 43697 MW; 31CAC841B6E5AC52 CRC64;

Query Match 45.7%; Score 37; DB 1; Length 396;
Best Local Similarity 40.0%; Pred. No. 63;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAPV 16
Db 82 EQKPLVLSALSRPII 96

RESULT 22
SDC3_MOUSE
ID SDC3_MOUSE STANDARD; PRT; 442 AA.
AC Q64519;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Syndecan-3 precursor (SYND3).
GN SDC3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA;
RX "Cloning of rat and mouse syndecan-3 cDNAs.";
RT Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RL -!- FUNCTION: CELL SURFACE PROTEOGLYCAN THAT MAY BEAR HEPARAN SULFATE.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PTM: O-GLYCOSYLATED WITHIN THE THR/SER-RICH REGION WHICH COULD
INTERACT WITH LECTIN DOMAINS ON OTHER MOLECULES (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE SYNDECAN FAMILY OF INTEGRAL MEMBRANE
PROTEOGLYCANS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U52826; AAB03283.1; -
DR MGD; MGI:1349163; Sdc3.
DR InterPro; IPR003585; Neurexin-like.
DR InterPro; IPR001050; Syndecan.
DR Pfam; PF01034; Syndecan; 1.
DR SMART; SM00294; 4.1m; 1.
DR PROSITE; PS00964; SYNDECAN; 1.
KW Proteoglycan; Heparan sulfate; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 45
FT CHAIN 46 442
FT DOMAIN 46 384
FT TRANSMEM 385 409
FT DOMAIN 410 442
FT DOMAIN 114 302
FT SITE 383 384
FT CARBOHYD 78 78
FT CARBOHYD 80 80
FT CARBOHYD 82 82
FT CARBOHYD 89 89
FT CARBOHYD 315 315
FT CARBOHYD 367 367
SQ SEQUENCE 442 AA; 46101 MW; D2F1C6B611FA164 CRC64;

Query Match 45.7%; Score 37; DB 1; Length 442;
Best Local Similarity 57.1%; Pred. No. 71;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 6 PLTTAAAPVVKNA 19
Db 216 PLTTAAATAKITTPA 229

RESULT 23
CATA_VIBFI
ID CATA_VIBFI STANDARD; PRT; 482 AA.
AC O68146;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Catalase (EC 1.11.1.6).
GN KATA.
```

```

OS  Vibrio fischeri.
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC  Vibrionaceae; Vibrio.
OX  NCBI_TaxID=668;
(1)
RN  SEQUENCE FROM N.A.
RP  STRAIN=ES114;
RC  MEDLINE=98215175; PubMed=9555890;
RX  Visick K., Ruby E.G.;
RT  "The periplasmic, group III catalase of Vibrio fischeri is required
RT  for normal symbiotic competence and is induced both by oxidative
RT  stress and by approach to stationary phase.";
RL  J. Bacteriol. 180:2087-2092(1998).
CC  -!- FUNCTION: Decomposes hydrogen peroxide into water and oxygen;
CC  serves to protect cells from the toxic effects of hydrogen
CC  peroxide. Could protect cells in nodules which have a high
CC  potential to produce hydrogen peroxide because of the strong
CC  reducing conditions required for nitrogen fixation and the action
CC  of several proteins.
CC  -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC  -!- COFACTOR: HEME GROUP.
CC  -!- SUBCELLULAR LOCATION: Periplasmic.
CC  -!- INDUCTION: By hydrogen peroxide.
CC  -!- MISCELLANEOUS: EITHER KATA OR KATC IS ABSOLUTELY REQUIRED FOR THE
CC  PROTECTION OF THE NITROGEN FIXATION PROCESS.
CC  -!- SIMILARITY: Belongs to the catalase family.
CC
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC
CC  EMBL; AF011784; AAC38344.1; -.
CC  HSP; P42321; 2CAE.
CC  InterPro; IPR002226; Catalase.
CC  Pfam; PF00199; catalase; 1.
CC  PRINTS; PR00067; CATALASE.
CC  ProDom; PD000510; Catalase; 1.
CC  ProSITE; PS00437; CATALASE_1; 1.
CC  PROSITE; PS00438; CATALASE_2; 1.
CC  Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
KW  Periplasmic.
FT  ACT_SITE 53 53 BY SIMILARITY.
FT  ACT_SITE 126 126 BY SIMILARITY.
FT  ACT_SITE 336 336 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT  METAL 336 336
SQ  SEQUENCE 482 AA; 54864 MW; 3808F0DBDE4DCC32 CRC64;

Query Match 45.7%; Score 37; DB 1; Length 482;
Best Local Similarity 66.7%; Pred. No. 77;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 7 LTTAAXAPVVXN 18
Db 5 LTTAAGCPVAHN 16

RESULT 24
LCYB CAPAN STANDARD; PRT; 498 AA.
AC Q43415;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lycopene beta cyclase, chloroplast precursor (EC 1.14.-.-).
GN LCY1 OR CRT1.
OS Capsicum annuum (Bell pepper).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiids; Solanales; Solanaceae; Capsicum.
OX NCBI_TaxID=4072;

```

```

(1)
RN  SEQUENCE FROM N.A.
RP  STRAIN=cv. Lamuyo; TISSUE=Fruit;
RX  MEDLINE=96045549; PubMed=7550379;
RA  Huguency P., Badillo A., Chen H.C., Klein A., Hirschberg J.,
RA  Camara B., Kuntz M.;
RT  "Metabolism of cyclic carotenoids: a model for the alteration of this
RT  biosynthetic pathway in Capsicum annum chromoplasts.";
RL  Plant J. 8:417-424(1995).
CC  -!- FUNCTION: CATALYZES THE DOUBLE CYCLIZATION REACTION WHICH CONVERTS
CC  LYCOPENE TO BETA-CAROTENE AND NEUROSPORENE TO BETA-ZEACAROTENE.
CC  -!- PATHWAY: Carotenoid biosynthesis.
CC  -!- SUBCELLULAR LOCATION: Chloroplast; chromoplast.
CC  -!- SIMILARITY: BELONGS TO THE LYCOPENE CYCLASE FAMILY.
CC
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC
CC  EMBL; X86221; CAA60119.1; -.
CC  InterPro; IPR001327; FAD pyr. redox.
CC  InterPro; IPR00205; NAD binding.
CC  InterPro; IPR001100; Pyr_redox.
CC  PRINTS; PR00368; FADPNR.
CC  PRINTS; PR00411; PNDRPTASE1.
CC  Oxidoreductase; NAD; Carotenoid biosynthesis; Chloroplast;
KW  Oxidoreductase; NAD; Carotenoid biosynthesis; Chloroplast;
KW  Transit peptide.
FT  TRANSIT 1 79 CHLOROPLAST (POTENTIAL).
FT  CHAIN 80 498 LYCOPENE BETA CYCLASE.
FT  NP_BIND 84 112 NAD (POTENTIAL).
SQ  SEQUENCE 498 AA; 55610 MW; 177180CD5745F64F CRC64;

Query Match 45.7%; Score 37; DB 1; Length 498;
Best Local Similarity 72.7%; Pred. No. 80;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 9 TAAAPVVXNA 19
Db 370 TLAAPVVANA 380

RESULT 25
LCYB LYCES
ID LCYB LYCES STANDARD; PRT; 500 AA.
AC Q43503;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lycopene beta cyclase, chloroplast precursor (EC 1.14.-.-).
GN LCY1 OR CRT1-1.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN  SEQUENCE FROM N.A.
RP  STRAIN=cv. VF36; TISSUE=Leaf;
RX  MEDLINE=96434545; PubMed=8837512;
RA  Cunningham F.X. Jr., Pogson B., Sun Z., McDonald K.A., Dellapenna D.,
RA  Gantt E.;
RT  "Functional analysis of the beta and epsilon lycopene cyclase enzymes
RT  of Arabidopsis reveals a mechanism for control of cyclic carotenoid
RT  formation.";
RL  Plant Cell 8:1613-1626(1996).
CC  -!- FUNCTION: CATALYZES THE DOUBLE CYCLIZATION REACTION WHICH CONVERTS
CC  LYCOPENE TO BETA-CAROTENE AND NEUROSPORENE TO BETA-ZEACAROTENE.
CC  -!- PATHWAY: Carotenoid biosynthesis.
CC  -!- SUBCELLULAR LOCATION: Chloroplast.

```

CC -!- SIMILARITY: BELONGS TO THE LYCOPENE CYCLASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X86452; CAA60170.1; -;
 CC PIR; S72505; S72505.
 CC InterPro; IPR001327; PAD_Pyr_redox.
 CC InterPro; IPR000205; NAD binding.
 CC InterPro; IPR001100; Pyr_redox.
 CC PRINTS; PR00368; FADPNR.
 CC PRINTS; PR00411; PNDRTASEI.
 CC Oxidoreductase; NAD; Carotenoid biosynthesis; Chloroplast;
 KW Transit peptide.
 FT TRANSIT 1 81 CHLOROPLAST (POTENTIAL).
 FT CHAIN 82 500 LYCOPENE BETA CYCLASE.
 FT NP BIND 86 114 NAD (POTENTIAL).
 SQ SEQUENCE 500 AA; 56180 MW; CF42F7D4684C04DD CRC64;
 Query Match 45.7%; Score 37; DB 1; Length 500;
 Best Local Similarity 72.7%; Pred. No. 80;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 9 TAAXAPVVXNA 19
 DB 372 TLAAAPVVANA 382
 RESULT 26
 LCYB_TOBAC STANDARD; PRT; 500 AA.
 AC Q43578;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Lycopene beta cyclase, chloroplast precursor (EC 1.14.-.-).
 GN LCY1 OR CRT1-1.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Samsun NN; TISSUE=Leaf;
 RX MEDLINE=96434545; PubMed=8837512;
 RA Cunningham F.X. Jr., Pogson B., Sun Z., McDonald K.A., Dellapenna D.,
 RA Gant E.;
 RA "Functional analysis of the beta and epsilon lycopene cyclase enzymes
 RT of Arabidopsis reveals a mechanism for control of cyclic carotenoid
 RT formation.";
 RL Plant Cell 8:1613-1626(1996).
 CC -!- FUNCTION: CATALYZES THE DOUBLE CYCLIZATION REACTION WHICH CONVERTS
 CC LYCOPENE TO BETA-CAROTENE AND NEUROSPORENE TO BETA-ZEAXANTHENE.
 CC -!- PATHWAY: Carotenoid biosynthesis.
 CC -!- SUBCELLULAR LOCATION: Chloroplast.
 CC -!- SIMILARITY: BELONGS TO THE LYCOPENE CYCLASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X81787; CAA57386.1; -;
 CC PIR; S72506; S72506.

DR InterPro; IPR000205; NAD binding.
 KW Oxidoreductase; NAD; Carotenoid biosynthesis; Chloroplast;
 FT TRANSIT 1 81 CHLOROPLAST (POTENTIAL).
 FT CHAIN 82 500 LYCOPENE BETA CYCLASE.
 FT NP BIND 86 114 NAD (POTENTIAL).
 SQ SEQUENCE 500 AA; 56067 MW; 2E3721B87EE6CBC8 CRC64;
 Query Match 45.7%; Score 37; DB 1; Length 500;
 Best Local Similarity 72.7%; Pred. No. 80;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 9 TAAXAPVVXNA 19
 DB 372 TLAAAPVVANA 382
 RESULT 27
 PSD3_ANOST STANDARD; PRT; 500 AA.
 ID PSD3 ANOST
 AC Q9U5Z8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable 26S proteasome non-ATPase regulatory subunit 3 (26S
 DE proteasome subunit S3) (Diphenol oxidase A2 component) (DOX-A2).
 GN DOXA2 OR DOX-A2.
 OS Anopheles stephensi (Indo-Pakistan malaria mosquito).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
 OX NCBI_TaxID=30069;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-StMal.
 RX MEDLINE=20296811; PubMed=10835480;
 RA Garvey C.F., Malcolm C.A.;
 RT "Anopheles stephensi Dox-A2 shares common ancestry with genes from
 RT distant groups of eukaryotes encoding a 26S proteasome subunit and is
 RT in a conserved gene cluster.";
 RL J. Mol. Evol. 50:497-509(2000).
 CC -!- FUNCTION: ACTS AS A REGULATORY SUBUNIT OF THE 26 PROTEASOME WHICH
 CC IS INVOLVED IN THE ATP-DEPENDENT DEGRADATION OF UBIQUITINATED
 CC PROTEINS (BY SIMILARITY).
 CC -!- SUBUNIT: THE 26S PROTEASOME IS COMPOSED OF A CORE PROTEASE, KNOWN
 CC AS THE 20S PROTEASOME, CAPPED AT ONE OR BOTH ENDS BY THE 19S
 CC REGULATORY COMPLEX (RC). THE RC IS COMPOSED OF AT LEAST 18
 CC DIFFERENT SUBUNITS IN TWO SUBCOMPLEXES, THE BASE AND THE LID,
 CC WHICH FORM THE PORTIONS PROXIMAL AND DISTAL TO THE 20S PROTEOLYTIC
 CC CORE, RESPECTIVELY (BY SIMILARITY).
 CC -!- DOMAIN: CONTAINS 1 PCI DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE PROTEASOME SUBUNIT S3 FAMILY.
 CC -!- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO BE THE DIPHENOL OXIDASE
 CC A2 COMPONENT INVOLVED IN CATECHOLAMINE METABOLISM, MELANIN
 CC FORMATION, AND SCLEROTIZATION OF THE CUTICLE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ250874; CAB61220.1; -;
 CC InterPro; IPR000717; PCI.
 CC Pfam; PF01399; PCI; 1.
 CC SMART; SM00088; PINT; 1.
 KW Proteasome.
 SQ SEQUENCE 500 AA; 56840 MW; 4B98EAO702EEB5C CRC64;
 Query Match 45.7%; Score 37; DB 1; Length 500;
 Best Local Similarity 58.3%; Pred. No. 80;
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```

QY 5 TPLTTAAXAPVV 16
DB 7 TPTATAASEPIV 18

RESULT 28
DHAY_YEAST STANDARD; PRT; 511 AA.
AC P32872;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aldehyde dehydrogenase 2, mitochondrial precursor (EC 1.2.1.3).
GN ALD2 OR ALDH2.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]_TaxID=4932;
RP SEQUENCE FROM N.A.
RL Thielen J.;
RL Thesis (1993), Heinrich-Heine University / Duesseldorf, Germany.
CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -!- PATHWAY: Ethanol utilization; second step.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix (Potential).
CC -!- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
CC -!- CAUTION: IS NOT PRESENT IN YEAST GENOME.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z17314; CAA78962.1; --
DR PIR; S31308; S31308.
DR HSP; P05091; ICW3.
DR SGD; L0000075; ALD2.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD; Mitochondrion; Transit peptide.
FT TRANSIT 1 21 MITOCHONDRION.
FT CHAIN 22 511 ALDEHYDE DEHYDROGENASE 2.
FT NP BIND 274 279 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 297 297 BY SIMILARITY.
FT ACT_SITE 331 331 BY SIMILARITY.
SQ SEQUENCE 511 AA; 56466 MW; 70EDAE951B94EB4A CRC64;

Query Match 45.7%; Score 37; DB 1; Length 511;
Best Local Similarity 44.4%; Pred. No. 82;
Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 EKTPLTTAAXAPVVXNA 19
DB 223 EELTPLTALTATVATLKEA 240

RESULT 29
YAJE_SCHPO STANDARD; PRT; 534 AA.
AC Q09911;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C30D11.14c in chromosome I.
GN SPAC30D11.14c.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

QY 2 EKTPLTTAAXAPVV 16
DB 415 QEKTOPTNASAPLV 429

Query Match 45.7%; Score 37; DB 1; Length 534;
Best Local Similarity 53.3%; Pred. No. 86;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKTPLTTAAXAPVV 16
DB 415 QEKTOPTNASAPLV 429

RESULT 30
MUC1_XENLA STANDARD; PRT; 662 AA.
ID MUC1_XENLA
AC Q05049;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Integumentary mucin C.1 (FIM-C.1) (Fragment).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]_TaxID=8355;

```

Schizosaccharomycetales; Schizosaccharomycetaceae;

Schizosaccharomycetes.

NCBI_TaxID=4896;

[1]_TaxID=4896;

SEQUENCE FROM N.A.

RP STRAIN=972;

RX MEDLINE=21848401; PubMed=11859360;

RA Wood V., Guillermo R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,

RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,

RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,

RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,

RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,

RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,

RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,

RA Rutherford J., Rutter S., Saunders D., Seeger K., Sharp S.,

RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,

RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,

RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,

RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,

RA Borzym K., Langer I., Beck A., Leirach H., Wambutt R., Purnelle B.,

RA Eger P., Zimmermann W., Wedler H., Gloux S., Lelaure V., Mottier S.,

RA Goffeau A., Cadieu E., Dreano S., Moore K., Hurst S.M.,

RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,

RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,

RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,

RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,

RA Shpakowski G.V., Ussery D., Barrell B.G., Nurse P.;

RT "The genome sequence of Schizosaccharomyces pombe.";

RL Nature 415:871-880(2002).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; Z67961; CAA91900.1; --

DR PIR; T38585; S62572.

DR GenDB SPombe; SPAC30D11.14c; --

DR InterPro; IPR004087; KH_dom.

DR SMART; SM00322; KH; 1.

DR Hypothetical protein.

SQ SEQUENCE 534 AA; 59035 MW; B38F20B8A20A47F7 CRC64;

Query Match 45.7%; Score 37; DB 1; Length 534;

Best Local Similarity 53.3%; Pred. No. 86;

Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKTPLTTAAXAPVV 16

DB 415 QEKTOPTNASAPLV 429

RESULT 30

MUC1_XENLA STANDARD; PRT; 662 AA.

ID MUC1_XENLA

AC Q05049;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Integumentary mucin C.1 (FIM-C.1) (Fragment).

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

OC Xenopodinae; Xenopus.

OX NCBI_TaxID=8355;

RN [1]_TaxID=8355;

SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4; 5; 6 AND 7).
TISSUE=Skin;
RC MEDLINE=93077556; PubMed=1447205;
RA Hauser F., Hoffmann W.;
RT "P-domains as shuffled cysteine-rich modules in integumentary mucin
C.1 (FIM-C.1) from *Xenopus laevis*. Polydispersity and genetic
polymorphism.";
RL J. Biol. Chem. 267:24620-24624(1992).
CC -!- FUNCTION: Could be involved in defense against microbial
infections. Protects the epithelia from external environment.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=7;
CC Comment=Additional isoforms seem to exist. Experimental
CC confirmation may be lacking for some isoforms;
CC Name=1;
CC IsoId=Q05049-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q05049-2; Sequence=VSP_004650;
CC Name=3;
CC IsoId=Q05049-3; Sequence=VSP_004651;
CC Name=4;
CC IsoId=Q05049-4; Sequence=VSP_004647, VSP_004648;
CC Name=5;
CC IsoId=Q05049-5; Sequence=VSP_004646, VSP_004649, VSP_004650;
CC Name=6;
CC IsoId=Q05049-6; Sequence=VSP_004646, VSP_004648;
CC Name=7;
CC IsoId=Q05049-7; Sequence=VSP_004647;
CC -!- TISSUE SPECIFICITY: SKIN.
CC -!- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC -!- SIMILARITY: Contains 6 P-type (trefoil) domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L02115; AAA74725.1; -
CC PIR; A45155; A45155.
CC HSP; P01359; 2PSP.
CC InterPro; IPR000519; P_trefoil.
CC Pfam; PF00088; trefoil; 6.
CC PRINTS; PR00680; PTREFOIL.
CC SMART; SM00018; PD; 6.
CC PROSITE; PS00025; P_TREFOIL; 6.
KW Repeat; Glycoprotein; Alternative splicing.
FT NON TER 1 1
FT DOMAIN 81 144 8 X 8 AA APPROXIMATE TANDEM REPEATS,
FT ALA/THR-RICH.
FT REPEAT 81 88
FT REPEAT 89 96 1-1.
FT REPEAT 97 104 1-2.
FT REPEAT 103 104 1-3.
FT REPEAT 105 112 1-4.
FT REPEAT 113 120 1-5.
FT REPEAT 121 128 1-6.
FT REPEAT 129 136 1-7.
FT REPEAT 137 144 1-8.
FT DOMAIN 161 202 P-TYPE 1.
FT DOMAIN 218 301 8 X APPROXIMATE TANDEM REPEATS, THR-RICH.
FT REPEAT 218 224 2-1.
FT REPEAT 225 239 2-2.
FT REPEAT 240 259 2-3.
FT REPEAT 250 259 2-4.
FT REPEAT 260 275 2-5.
FT REPEAT 276 287 2-6.
FT REPEAT 288 294 2-7.
FT REPEAT 295 301 2-8.
FT DOMAIN 306 347 P-TYPE 2.
FT DOMAIN 353 394 P-TYPE 3.

FT DOMAIN 402 522 12 X APPROXIMATE TANDEM REPEATS, THR-
FT RICH.
FT REPEAT 402 411 3-1.
FT REPEAT 412 419 3-2.
FT REPEAT 420 431 3-3.
FT REPEAT 432 443 3-4.
FT REPEAT 444 453 3-5.
FT REPEAT 454 460 3-6.
FT REPEAT 461 472 3-7.
FT REPEAT 473 479 3-8.
FT REPEAT 480 491 3-9.
FT REPEAT 492 498 3-10.
FT REPEAT 499 515 3-11.
FT REPEAT 516 522 3-12.
FT DOMAIN 525 566 P-TYPE 4.
FT DOMAIN 572 613 P-TYPE 5.
FT DOMAIN 620 661 P-TYPE 6.
FT DISULFID 162 188 BY SIMILARITY.
FT DISULFID 172 187 BY SIMILARITY.
FT DISULFID 182 199 BY SIMILARITY.
FT DISULFID 307 333 BY SIMILARITY.
FT DISULFID 317 332 BY SIMILARITY.
FT DISULFID 327 344 BY SIMILARITY.
FT DISULFID 354 380 BY SIMILARITY.
FT DISULFID 374 379 BY SIMILARITY.
FT DISULFID 526 552 BY SIMILARITY.
FT DISULFID 536 551 BY SIMILARITY.
FT DISULFID 546 563 BY SIMILARITY.
FT DISULFID 573 599 BY SIMILARITY.
FT DISULFID 583 598 BY SIMILARITY.
FT DISULFID 593 610 BY SIMILARITY.
FT DISULFID 621 647 BY SIMILARITY.
FT DISULFID 631 646 BY SIMILARITY.
FT DISULFID 641 658 BY SIMILARITY.
FT VARSPLIC 240 259 Missing (in isoform 5 and isoform 6).
FT VARSPLIC 250 259 Missing (in isoform 4 and isoform 7).
FT VARSPLIC 276 294 Missing (in isoform 4 and isoform 6).
FT VARSPLIC 278 278 Missing (in isoform 5).
FT VARSPLIC 306 350 Missing (in isoform 2 and isoform 5).
FT VARSPLIC 420 498 Missing (in isoform 3).
FT VARIANT 276 276 K -> E.
FT VARIANT 354 354 C -> R.
FT VARIANT 415 415 T -> A.
SQ SEQUENCE 662 AA; 67774 MW; F085277F1ED2FD40 CRC64;
Query Match 45.7%; Score 37; DB 1; Length 662;
Best Local Similarity 50.0%; Pred. No. 1.le+02;
Matches 8; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 4 KTEPLTTAAKAPVXNA 19
DB 81 KAPTAAATAPTAAA 96
RESULT 31
FTSK CAMJE STANDARD; PRT; 946 AA.
ID FTSK CAMJE Q9PP45;
AC Q46089; Q9PP45; 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cell division protein ftsk homolog.
GN FTSK OR CJ0886C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.

DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Serine protease inhibitor Kazal-type 5 precursor (Lympho-epithelial
DE Kazal-type related inhibitor) (LEKTI) [Contains: Hemofiltrate peptide
DE HF6478; Hemofiltrate peptide HF7665].
GN SPINK5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND VARIANT GLU-420.
RC TISSUE=Epithelium;
RX MEDLINE=99348267; PubMed=10419450;
RA Maegert H.-J., Staendker L., Kreutzmann P., Zucht H.-D., Reinecke M.,
RA Sommerhoff C.P., Fritz H., Forstmann W.-G.;
RT "LEKTI, a novel 15-domain type of human serine proteinase inhibitor.";
RL J. Biol. Chem. 274:21499-21502(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20296612; PubMed=10835624;
RA Chavanas S., Bodemer C., Rochat A., Hamel-Teillac D., Ali M.,
RA Irvine A.D., Bonafe J.-L., Wilkinson J., Taieb A., Barrandon Y.,
RA Harper J.I., de Prost Y., Hovnanian A.;
RT "Mutations in SPINK5, encoding a serine protease inhibitor, cause
RT Netherton syndrome.";
RL Nat. Genet. 25:141-142(2000).
RN [3]
RP SEQUENCE OF 202-222 AND 266-294 FROM N.A.
RX MEDLINE=21403086; PubMed=11511232;
RA Sprecher E., Chavanas S., DiGiovanna J.J., Amin S., Nielsen K.,
RA Prendiville J.S., Silverman R., Esterly N.B., Spraker M.K., Guelig E.,
RA de Luna M.L., Williams M.L., Buehler B., Siegfried B.C.,
RA Van Maldergem L., Pfendner E., Bale S.J., Uitto J., Hovnanian A.,
RA Richard G.;
RT "The spectrum of pathogenic mutations in SPINK5 in 19 families with
RT Netherton syndrome: Implications for mutation detection and first case
RT of prenatal diagnosis.";
RL J. Invest. Dermatol. 117:179-187(2001).
RN [4]
RP SEQUENCE OF 490-507.
RX TISSUE=Foreskin keratinocyte;
RX MEDLINE=21477652; PubMed=11594460;
RA Ahmed A., Kandola P., Ziada G., Parenteau N.;
RA "Purification and partial amino acid sequence of proteins from human
RT epidermal keratinocyte conditioned medium.";
RL J. Protein Chem. 20:273-278(2001).
RN [5]
RP VARIANT GLU-420.
RA Walley A.J., Chavanas S., Moffatt M.F., Esmouf R.M., Ubhi B.,
RA Lawrence R., Wong K., Abecasis G.R., Jones E.Y., Harper J.I.,
RA Hovnanian A., Cookson W.O.C.M.;
RT "Gene polymorphism in Netherton and common atopic disease.";
RL Nat. Genet. 29:175-178(2001).
CC -!- FUNCTION: SERINE PROTEASE INHIBITOR, PROBABLY IMPORTANT FOR THE
CC ANTI-INFLAMMATORY AND/OR ANTIMICROBIAL PROTECTION OF MUCOUS
CC EPITHELIA.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE THYMUS. ALSO FOUND IN
CC THE ORAL MUCOSA, PARATHYROID GLAND, BARTHOLIN'S GLANDS, TONSILS,
CC AND VAGINAL EPITHELIUM. VERY LOW LEVELS ARE DETECTED IN LUNG,
CC KIDNEY, AND PROSTATE.
CC -!- DOMAIN: CONTAINS AT LEAST ONE ACTIVE INHIBITORY DOMAIN FOR TRYPSIN
CC (DOMAIN 6).
CC -!- DISEASE: Defects in SPINK5 are the cause of Netherton syndrome
CC (NS) [MIM:256500], a severe autosomal recessive disorder. It is
CC characterized by congenital ichthyosis with defective
CC cornification, a specific hair shaft defect known as trichorrhexis
CC invaginata or "bamboo hair", atopic dermatitis, and hayfever. High
CC post-natal mortality is due to failure to thrive, infections and
CC hypernatraemic dehydration.
CC -!- SIMILARITY: Contains 15 Kazal-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC -----

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ228139; CAB40839.1; -
DR EMBL; AJ391230; CAB96877.1; JOINED.
DR EMBL; AJ270944; CAB96877.1; JOINED.
DR EMBL; AJ391231; CAB96877.1; JOINED.
DR EMBL; AJ391232; CAB96877.1; JOINED.
DR EMBL; AJ391233; CAB96877.1; JOINED.
DR EMBL; AJ391234; CAB96877.1; JOINED.
DR EMBL; AJ391235; CAB96877.1; JOINED.
DR EMBL; AJ276579; CAB96877.1; JOINED.
DR EMBL; AJ391236; CAB96877.1; JOINED.
DR EMBL; AJ276580; CAB96877.1; JOINED.
DR EMBL; AJ391237; CAB96877.1; JOINED.
DR EMBL; AJ391238; CAB96877.1; JOINED.
DR EMBL; AJ391239; CAB96877.1; JOINED.
DR EMBL; AJ391240; CAB96877.1; JOINED.
DR EMBL; AJ391241; CAB96877.1; JOINED.
DR EMBL; AJ276578; CAB96877.1; JOINED.
DR EMBL; AJ391242; CAB96877.1; JOINED.
DR EMBL; AJ391243; CAB96877.1; JOINED.
DR EMBL; AJ391244; CAB96877.1; JOINED.
DR EMBL; AJ391245; CAB96877.1; JOINED.
DR EMBL; AJ391246; CAB96877.1; JOINED.
DR EMBL; AJ391247; CAB96877.1; JOINED.
DR EMBL; AJ391248; CAB96877.1; JOINED.
DR EMBL; AJ391249; CAB96877.1; JOINED.
DR EMBL; AJ391250; CAB96877.1; JOINED.
DR EMBL; AJ391251; CAB96877.1; JOINED.
DR EMBL; AJ391252; CAB96877.1; JOINED.
DR EMBL; AJ391253; CAB96877.1; JOINED.
DR EMBL; AJ391254; CAB96877.1; JOINED.
DR EMBL; AJ276577; CAB96877.1; JOINED.
DR EMBL; AF295784; AAK97139.1; -
DR HSSP; AF295783; AAK97140.1; -
DR Genew; HGNC:15464; SPINK5.
DR MIM; 605010; -
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0004867; F:serine protease inhibitor activity; NAS.
DR GO; GO:0003026; P:anti-inflammatory response; NAS.
DR GO; GO:0006960; P:antimicrobial humoral response (sensu Inver. . .); NAS.
DR InterPro; IPR002350; kazal.
DR Pfam; PF00050; kazal; 7.
DR SMART; SM00280; kazal; 14.
DR PROSITE; PS00282; KAZAL; 2.
KW Serine protease inhibitor; Repeat; Signal; Polymorphism.
FT SIGNAL 1 22
FT CHAIN 23 1064
FT PEPTIDE 23 77
FT PEPTIDE 356 423
FT DOMAIN 28 66
FT DOMAIN 95 151
FT DOMAIN 159 214
FT DOMAIN 223 272
FT DOMAIN 295 351
FT DOMAIN 365 421
FT DOMAIN 435 473
FT DOMAIN 494 556
FT DOMAIN 562 621
FT DOMAIN 630 686
FT DOMAIN 705 757
FT DOMAIN 773 828
FT DOMAIN 847 903
FT DOMAIN 914 970
FT DOMAIN 991 1046
FT DISULFID 30 66
BY SIMILARITY.

FT	CARBOHYD	2300	2300	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2304	2304	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2456	2456	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2702	2702	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2712	2712	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	3390 AA; 378057 MW; 665887FOFIE1756E CRC64;		
Query Match 45.7%; Score 37; DB 1; Length 3390;				
Best Local Similarity 53.8%; Pred. No. 5.7e+02;				
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;				
QY	2	EKTEPLTTAAAXAP 14		
DB	3315	EDKTEVTTWEDVP 3327		
RESULT 36				
CAVI_CABEL	STANDARD;	PRT;	235 AA.	
ID	CAVI_CABEL			
AC	Q94051;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	Caveolin-1.			
GN	CAV-1 OR T13F2.8.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;			
OC	Rhabditidae; Peloderinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Bristol N2;			
RX	MEDLINE=97153022; PubMed=8999956;			
RA	Tang Z, Okamoto T., Boontrakulpoontawe P., Katada T., Otsuka A.J.,			
RA	Lisanti M.P.;			
RT	"Identification, sequence, and expression of an invertebrate caveolin			
RT	gene family from the nematode Caenorhabditis elegans. Implications.			
RT	for the molecular evolution of mammalian caveolin genes.";			
RN	J. Biol. Chem. 272:2437-2445(1997).			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Bristol N2;			
RA	Swinburne J.;			
RL	Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.			
CC	!- FUNCTION: MAY ACT AS A SCAFFOLDING PROTEIN WITHIN CAVEOLAR			
CC	MEMBRANES. INTERACTS DIRECTLY WITH G-PROTEIN ALPHA SUBUNITS AND			
CC	CAN FUNCTIONALLY REGULATE THEIR ACTIVITY.			
CC	!- SUBUNIT: HOMOLIGOMER CONTAINING 14-16 MONOMERS PER OLIGOMER.			
CC	!- SUBCELLULAR LOCATION: MEMBRANE PROTEIN OF CAVEOLAE. POTENTIAL			
CC	HAIRPIN-LIKE STRUCTURE IN THE MEMBRANE (BY SIMILARITY).			
CC	!- SIMILARITY: BELONGS TO THE CAVEOLIN FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	EMBL; U66405; AAB48388.1; -			
CC	EMBL; 281122; CAB03359.1; -			
DR	PIR; T24882; T24882.			
DR	WormPep; T13F2.8; Cei13633.			
DR	InterPro; IPR001612; Caveolin.			
DR	Pfam; PF01146; Caveolin; 1.			
DR	PROSITE; PS01210; CAVEOLIN; FALSE NEG.			
KW	Transmembrane; Lipoprotein; Palmitate.			
FT	DOMAIN 1 161			
FT	TRANSMEM 162 182			
FT	DOMAIN 183 235			
FT	LIPID 234 234			
SQ	SEQUENCE 235 AA; 26291 MW; F07B12DEB4D6F13A CRC64;			

FT	CARBOHYD	2300	2300	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2304	2304	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2456	2456	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2702	2702	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2712	2712	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	3390 AA; 378057 MW; 665857FOFIE1756E CRC64;		
Query Match 45.7%; Score 37; DB 1; Length 3390;				
Best Local Similarity 53.8%; Pred. No. 5.7e+02;				
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;				
QY	2	EKTEPLTTAAAXAP 14		
DB	3315	EDKTEVTTWEDVP 3327		
RESULT 36				
CAVI_CABEL	STANDARD;	PRT;	235 AA.	
ID	CAVI_CABEL			
AC	Q94051;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	Caveolin-1.			
GN	CAV-1 OR T13F2.8.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;			
OC	Rhabditidae; Peloderinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Bristol N2;			
RX	MEDLINE=97153022; PubMed=8999956;			
RA	Tang Z, Okamoto T., Boontrakulpoontawe P., Katada T., Otsuka A.J.,			
RA	Lisanti M.P.;			
RT	"Identification, sequence, and expression of an invertebrate caveolin			
RT	gene family from the nematode Caenorhabditis elegans. Implications			
RT	for the molecular evolution of mammalian caveolin genes.";			
RN	J. Biol. Chem. 272:2437-2445(1997).			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Bristol N2;			
RA	Swinburne J.;			
RL	Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.			
CC	!- FUNCTION: MAY ACT AS A SCAFFOLDING PROTEIN WITHIN CAVEOLAR			
CC	MEMBRANES. INTERACTS DIRECTLY WITH G-PROTEIN ALPHA SUBUNITS AND			
CC	CAN FUNCTIONALLY REGULATE THEIR ACTIVITY.			
CC	!- SUBUNIT: HOMOLIGOMER CONTAINING 14-16 MONOMERS PER OLIGOMER.			
CC	!- SUBCELLULAR LOCATION: MEMBRANE PROTEIN OF CAVEOLAE. POTENTIAL			
CC	HAIRPIN-LIKE STRUCTURE IN THE MEMBRANE (BY SIMILARITY).			
CC	!- SIMILARITY: BELONGS TO THE CAVEOLIN FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	EMBL; U66405; AAB48388.1; -			
CC	EMBL; 281122; CAB03359.1; -			
DR	PIR; T24882; T24882.			
DR	WormPep; T13F2.8; Cei13633.			
DR	InterPro; IPR001612; Caveolin.			
DR	Pfam; PF01146; Caveolin; 1.			
DR	PROSITE; PS01210; CAVEOLIN; FALSE NEG.			
KW	Transmembrane; Lipoprotein; Palmitate.			
FT	DOMAIN 1 161			
FT	TRANSMEM 162 182			
FT	DOMAIN 183 235			
FT	LIPID 234 234			
SQ	SEQUENCE 235 AA; 26291 MW; F07B12DEB4D6F13A CRC64;			

```
Query Match 45.1%; Score 36.5; DB 1; Length 235;
Best Local Similarity 62.5%; Pred. No. 45;
Matches 10; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 2 EKTPLTTAA-XAPVV 16
   ||| ||| |||
DB 10 EEQIPLTYAAVAAPT 25

RESULT 37
HUNB DROMM STANDARD; PRT; 158 AA.
AC 046248; 046249;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hunchback protein (Fragments).
GN HB.
OS Drosophila mimica (Fruit fly) (Idiomysia mimica).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7270;
RN [1]
RP SEQUENCE FROM N.A.
RA Baker R.H., Desalle R.;
RT "Multiple sources of character information and the phylogeny of
RT Hawaiian Drosophilids.";
RL Syst. Biol. 46:654-673(1997).
CC -!- FUNCTION: GAP CLASS SEGMENTATION PROTEIN THAT CONTROLS DEVELOPMENT
CC OF HEAD STRUCTURES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: BELONGS TO THE HUNCHBACK FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U93012; AAC03260.1; --
CC FLYBASE; U93013; AAC03261.1; --
CC FlyBase; FBgn0023764; Dmhc\hb.
CC Developmental protein; Gap protein; Zinc-finger;
KW Metal-binding; DNA-binding; Repeat; Nuclear protein.
FT NON_TER 1
FT DOMAIN 17 31 POLY-HIS.
FT DOMAIN 60 64 POLY-GLN.
FT NON_CONS 64 65
FT DOMAIN 91 95 POLY-THR.
FT DOMAIN 96 99 POLY-ALA.
FT NON_TER 158 158
SQ SEQUENCE 158 AA; 17576 MW; 114B650BD4DC8CDE CRC64;

Query Match 44.4%; Score 36; DB 1; Length 158;
Best Local Similarity 50.0%; Pred. No. 37;
Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 4 KTLPLTTAAAPVXNA 19
   : ||| |||
DB 90 RTITTTAAAPTTAA 105

RESULT 38
DAMX_SERMA STANDARD; PRT; 214 AA.
AC P45459;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
```

```
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE DamX protein (Fragment).
GN DAMX.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Serratia.
OX NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Sr41;
RA Ostendorf T., Cherepanov P., Jekel M., de Vries J., Wackernagel W.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBSJ databases.
CC -!- FUNCTION: DIRECTLY OR INDIRECTLY INTERFERES WITH CELL DIVISION.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X78412; CAA55176.1; --
CC FIR; S47098; S47098.
FT NON_TER 1
SQ SEQUENCE 214 AA; 22157 MW; 3D1BDD43F8D6C401 CRC64;

Query Match 44.4%; Score 36; DB 1; Length 214;
Best Local Similarity 53.8%; Pred. No. 50;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAP 14
   : ||| |||
DB 79 QHKTAKTAAAKP 91

RESULT 39
ANT_BPPI STANDARD; PRT; 322 AA.
AC P19555;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Antirepressor protein 1 (ANT1) (Repressor bypass protein B) (REBB)
DE [Contains: Antirepressor protein 2 (ANT2) (Repressor bypass protein A)
DE (REBA)].
DE ANT OR REB.
GN Bacteriophage P1.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC P1-like viruses.
OX NCBI_TaxID=10678;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89293847; PubMed=2661831;
RA Hansen E.B.;
RT "Structure and regulation of the lytic replicon of phage P1.";
RL J. Mol. Biol. 207:135-149(1989).
CC -!- FUNCTION: Antagonist of the C1 repressor or repressor bypass
CC function.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X15639; CAA33659.1; --
CC FIR; S04261; ATBPP1.
CC InterPro; IPR005039; Anti_rep.
CC Pfam; PF03374; ANT; 1.
CC DNA-binding.
KW
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AL035675; CAB38691.2; -
 DR PIR; T39362; T39362.
 DR GeneDB.SPombe; SPBC1289.11; -
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 7.
 DR PRINTS; PR00320; GPROTEINBRPT.
 DR ProDom; PD000018; WD40; 2.
 DR SMART; SM00320; WD40; 7.
 DR PROSITE; PS00678; WD_REPEATS_1; 2.
 DR PROSITE; PS50082; WD_REPEATS_2; 7.
 DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
 KW Nuclear protein; mRNA processing; Repeat; WD repeat.
 FT REPEAT 48 87 WD 1.
 FT REPEAT 91 130 WD 2.
 FT REPEAT 133 173 WD 3.
 FT REPEAT 175 214 WD 4.
 FT REPEAT 217 256 WD 5.
 FT REPEAT 267 306 WD 6.
 FT REPEAT 308 339 WD 7.
 SQ SEQUENCE 340 AA; 37427 MW; BBAA5938CA4C24C2 CRC64;

Query Match 44.4%; Score 36; DB 1; Length 322;
 Best Local Similarity 46.7%; Pred. No. 76;
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 ECKTPLTTAAAPV 16
 | : : : : : |
 Db 92 EKPATVTSATPAP 106

RESULT 40

CWFF SCHPO STANDARD; PRT; 340 AA.
 AC O94620;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cell cycle control protein cwf17.
 GN CWF17 OR SPBC1289.11.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Grynoprez B.,
 RA Welltjens I., Vanstreels E., Rieger M., Schaefer M., Muelier-Auer S.,
 RA Borzym K., Langer I., Beck A., Holzer C., Holzner H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe."
 RL Nature 415:871-880(2002).
 RN [2]
 RP IDENTIFICATION.
 RX MEDLINE=21881936; PubMed=11884590;
 RA Ohi M.D., Link A.J., Ren L., Jennings J.L., McDonald W.H., Gould K.L.;
 RA "Proteomics analysis reveals stable multiprotein complexes in both
 RT fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel
 RT pre-mRNA splicing factors, and snRNAs."
 RL Mol. Cell. Biol. 22:2011-2024(2002).
 CC -!- FUNCTION: Involved in mRNA splicing where it associates with cdc5
 CC and the other cwf proteins as part of the spliceosome.
 CC -!- SUBUNIT: Interacts with cdc5
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- SIMILARITY: Contains 7 WD repeats.
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration

Query Match 44.4%; Score 36; DB 1; Length 340;
 Best Local Similarity 46.7%; Pred. No. 80;
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ECKTPLTTAAAXA 13
 | : : : : : |
 Db 175 EKYPLTAVATA 186

Search completed: January 28, 2004, 13:08:34
 Job time : 12 secs

THIS PAGE BLANK (user to)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 13:05:59 ; Search time 34 Seconds
(without alignments)
144.206 Million cell updates/sec

Title: US-09-359-426C-2

Perfect score: 81

Sequence: 1 XEEKTLTAAAXAPVVXNA 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23.*

- 1: sp archaea.*
- 2: sp bacteria.*
- 3: sp fungi.*
- 4: sp human.*
- 5: sp invertebrate.*
- 6: sp mammal.*
- 7: sp mhc.*
- 8: sp organelle.*
- 9: sp phase.*
- 10: sp plant.*
- 11: sp rodent.*
- 12: sp virus.*
- 13: sp vertebrate.*
- 14: sp unclassified.*
- 15: sp rvirus.*
- 16: sp bacteriap.*
- 17: sp archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	61.7	485	2 Q8GRA4	Q8gra4 porphyromon
2	48	59.3	480	16 Q8ZGS4	Q8zgs4 yersinia pe
3	46	56.8	458	5 O76672	O76672 caenorhabdi
4	45	55.6	484	2 Q9RG14	Q9rg14 actinobacil
5	44	54.3	306	16 Q8XGW5	Q8xgw5 salmonella
6	44	54.3	306	16 Q8PKR4	Q8pkr4 escherichia
7	44	54.3	314	2 Q9A119	Q9a119 escherichia
8	44	54.3	427	10 Q94AQ0	Q94aq0 arabidopsis
9	44	54.3	427	10 Q23188	Q23188 arabidopsis
10	43	53.1	486	2 Q8KKH4	Q8kqh4 hyphomicrob
11	43	53.1	506	2 P77924	P77924 pseudomonas
12	43	53.1	1240	12 Q9DWH8	Q9dwh8 rat cytomeg
13	42	51.9	210	2 Q9RNU2	Q9rnu2 streptococc
14	42	51.9	211	2 Q9X6X8	Q9x6x8 streptococc
15	42	51.9	436	10 Q9FT45	Q9ft45 arabidopsis
16	42	51.9	487	16 Q9RJK9	Q9rjk9 streptomyce

17	41	50.6	132	10 Q9SE04	Q9se04 oryza sativ
18	41	50.6	417	2 Q54272	Q54272 streptomyce
19	41	50.6	516	16 Q8P7W8	Q8p7w8 xanthomonas
20	41	50.6	663	17 Q8ZZX0	Q8zzx0 pyrobaculum
21	41	50.6	920	5 Q8MT63	Q8mt63 drosophila
22	41	50.6	920	5 Q8ML93	Q8ml93 drosophila
23	40	49.4	171	5 Q9U3E3	Q9u3e3 caenorhabdi
24	40	49.4	212	5 Q9U4Y3	Q9u4y3 ceratitis c
25	40	49.4	238	16 Q8PD46	Q8pd46 xanthomonas
26	40	49.4	238	16 Q8PD69	Q8pd69 xanthomonas
27	40	49.4	252	17 Q9Y984	Q9y984 aeropyrum p
28	40	49.4	258	16 Q8YSM9	Q8ysm9 listeria mo
29	40	49.4	367	5 Q8I1A1	Q8i1a1 drosophila
30	40	49.4	394	16 Q9A4B0	Q9a4b0 caulobacter
31	40	49.4	477	16 Q9RD97	Q9rd97 streptomyce
32	40	49.4	498	15 Q79665	Q79665 human immun
33	40	49.4	594	5 Q8BLH5	Q8blh5 halocynthia
34	40	49.4	788	5 Q9W180	Q9w180 drosophila
35	40	49.4	805	5 Q8MLP2	Q8mlp2 drosophila
36	40	49.4	846	13 Q57577	Q57577 cynops pyrr
37	40	49.4	1266	11 Q8CGF1	Q8cgf1 mus musculu
38	40	49.4	1360	5 Q02006	Q02006 drosophila
39	40	49.4	1399	5 Q9VHP9	Q9vhp9 drosophila
40	39.5	48.8	405	10 Q03462	Q03462 zea mays (m
41	39.5	48.8	5412	5 Q9W596	Q9w596 drosophila
42	39	48.1	209	12 Q8B8M6	Q8b8m6 human adeno
43	39	48.1	233	16 Q8NP51	Q8np51 corynebacte
44	39	48.1	255	10 Q9LGM5	Q9lgm5 oryza sativ
45	39	48.1	286	16 Q8P4V3	Q8p4v3 xanthomonas

ALIGNMENTS

RESULT 1
Q8GRA4 ID Q8GRA4 PRELIMINARY; PRT; 485 AA.
AC Q8GRA4;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Catalase HP2.
GN KAT.
OS Porphyromonas gulae.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=111105;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPB3492;
RA Nakayama K., Sato K.;
RT "catalase";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB083039; BAC20190.1;
SQ SEQUENCE 485 AA; 55610 MW; 6D35EC7A7E59F8AF CRC64;

Query Match 61.7%; Score 50; DB 2; Length 485;
Best Local Similarity 68.8%; Pred. No. 2.9;
Matches 11; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 EKTPLTTAAAXAPVVXN 18
||| ||||| |||
DB 2 EKNKLTAAAPVADN 17

RESULT 2
Q8ZGS4 ID Q8ZGS4 PRELIMINARY; PRT; 480 AA.
AC Q8ZGS4;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Catalase (EC 1.11.1.6).

```

GN KATA OR YPO1207 OR KATE OR Y2981.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parhill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holtroyd S., Jagels K., Kariyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RA "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
RN [2]
RN RP SEQUENCE FROM N.A.
RC STRAIN=KIMS / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RA "Genome sequence of Yersinia pestis KIM."
RT J. Bacteriol. 184:4601-4611(2002).
DR EMBL: AJ414147; CAC90045.1; -
DR EMBL: AE013900; AM86532.1; -
DR HSP: P42321; IM85.
DR InterPro: IPR002226; Catalase.
DR Pfam: PF00199; catalase; 1.
DR PRINTS: PR00067; CATALASE.
DR ProDom: PD000510; Catalase; 1.
DR PROSITE: PS00438; CATALASE_2; 1.
KW Oxidoreductase; Peroxidase; Complete proteome.
SQ SEQUENCE 480 AA; 55023 MW; 27A72FBAF096347E CRC64;

Query Match 59.3%; Score 48; DB 16; Length 480;
Best Local Similarity 68.8%; Pred. No. 6.3;
Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 EKTPLTTAAXAPVVXN 18
: ||||| |||||
Db 4 KKKGLTTAAGAPVVDN 19

RESULT 3
O76672 ID O76672 PRELIMINARY; PRT; 458 AA.
AC DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 47.3 kDa protein.
GN H34124.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RN RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;

```

```

RA Latreille P., Wamsley P., O'Brien D.;
RT "The sequence of C. elegans cosmid H34124."
RN Submitted (JUL-1998) to the EMBL/GenBank/DBSJ databases.
RL [3]
RN RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL: AF078784; AAKJ8967.1; -
DR WormPep: H34124.2; CE26681.
KW Hypothetical protein.
SQ SEQUENCE 458 AA; 47285 MW; 2AF5BC94EDA0C324 CRC64;

Query Match 56.8%; Score 46; DB 5; Length 458;
Best Local Similarity 56.2%; Pred. No. 13;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 KIPLTTAAXAPVVXNA 19
||||| : |||
Db 386 KIPLTSGSSARVINNA 401

RESULT 4
Q9RG14 ID Q9RG14 PRELIMINARY; PRT; 484 AA.
AC Q9RG14;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Catalase (EC 1.11.1.6).
GN KATA.
OS Actinobacillus actinomycetemcomitans (Haemophilus
OS actinomycetemcomitans).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Actinobacillus.
OX NCBI_TaxID=714;
RN [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=Y4Na1;
RX MEDLINE=20042351; PubMed=10572134;
RA Thomson V.J., Bhattacharjee M.K., Fine D.H., Derbyshire K.M.,
RA Figurski D.H.;
RT "Direct Selection of IS903 Transposon Insertions by Use of a Broad-
RT Host-Range Vector: Isolation of Catalase-Deficient Mutants of
RT Actinobacillus actinomycetemcomitans."
RL J. Bacteriol. 181:7298-7307(1999).
CC -1- FUNCTION: OCCURS IN ALMOST ALL AEROBICALLY RESPIRING ORGANISMS AND
CC SERVES TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN
CC PEROXIDE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
DR EMBL: AF162654; AAF17882.1; -
DR HSP: P42321; IM85.
DR InterPro: IPR002226; Catalase.
DR Pfam: PF00199; catalase; 1.
DR PRINTS: PR00067; CATALASE.
DR ProDom: PD000510; Catalase; 1.
DR PROSITE: PS00437; CATALASE_1; 1.
DR PROSITE: PS00438; CATALASE_2; 1.
KW Heme; Hydrogen peroxide; Iron; Oxidoreductase; Peroxidase.
SQ SEQUENCE 484 AA; 54961 MW; D0E523AB2557D8CB CRC64;

Query Match 55.6%; Score 45; DB 2; Length 484;
Best Local Similarity 83.3%; Pred. No. 21;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 LTITAAXAPVVXN 18
||||| |||||
Db 14 LTITAAGAPVVDN 25

RESULT 5
Q8XGW5 ID Q8XGW5 PRELIMINARY; PRT; 306 AA.

```

```
AC Q8XGW5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative sugar kinase, ribokinase family (EC 2.7.1.15) (Putative
DE carbohydrate kinase).
GN STM3793 OR STY3989.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601;
RN
RN
SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN
RN
SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL; AE008876; AAL23651.1; -
DR EMBL; AL627280; CAD03201.1; -
DR InterPro; IPR002173; PfKB.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; pfKB; 1.
DR PRINTS; PR00990; RIBOKINASE.
KW Kinase; Transferase; Hypothetical protein; Complete proteome.
SQ SEQUENCE 306 AA; 3329 MW; 142A0289E2D1B953 CRC64;

Query Match 54.3%; Score 44; DB 16; Length 306;
Best Local Similarity 47.1%; Pred. No. 19;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 EKTPLTTAXAPVXNA 19
DB 84 EKVPCTSSGVAPIFVNA 100
|||:::|:|
|||:::|:|

RESULT 6
Q8FKR4 PRELIMINARY; PRT; 306 AA.
ID Q8FKR4;
AC Q8FKR4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative ribokinase.
GN C0331.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN
RN
SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX
```

```
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
KW EMBL; AE016756; AAN78819.1; -
KW Kinase; Complete proteome.
SQ SEQUENCE 306 AA; 3322 MW; 8A0FBEP6BFE493DB CRC64;

Query Match 54.3%; Score 44; DB 16; Length 306;
Best Local Similarity 47.1%; Pred. No. 19;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 EKTPLTTAXAPVXNA 19
DB 84 EKVPCTSSGVAPIFVNA 100
|||:::|:|
|||:::|:|

RESULT 7
Q9A119 PRELIMINARY; PRT; 314 AA.
ID Q9A119;
AC Q9A119;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 34.4 kDa protein.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN
RN
SEQUENCE FROM N.A.
RC STRAIN=AL862;
RX MEDLINE=21101044; PubMed=11159989;
RA Lalioui L., Le Bouguenec C.C.;
RT "afa-8 Gene cluster is carried by a pathogenicity island inserted into
RT the tRNA(Phe) of human and bovine pathogenic Escherichia coli
RT isolates.";
RL Infect. Immun. 69:937-948(2001).
DR EMBL; AF286671; RAK27331.1; -
DR HSSP; P05054; LRK2.
DR InterPro; IPR002173; PfKB.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; pfKB; 1.
DR PRINTS; PR00990; RIBOKINASE.
KW Hypothetical protein.
SQ SEQUENCE 314 AA; 34415 MW; 2D5F7342C817E3BB CRC64;

Query Match 54.3%; Score 44; DB 2; Length 314;
Best Local Similarity 47.1%; Pred. No. 20;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 EKTPLTTAXAPVXNA 19
DB 93 EKVPCTSSGVAPIFVNA 109
|||:::|:|
|||:::|:|

RESULT 8
Q94AQ0 PRELIMINARY; PRT; 427 AA.
ID Q94AQ0;
AC Q94AQ0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical 46.8 kDa protein.
GN C7A10.390.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
```

```
OX NCBI_TaxID=3702;
RN [1] :||| :||| :|||
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Goldsmith A.D.,
RA Jiang P.X., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M.,
RA Yamamura Y., Yu G., Bowser L., Carninci P., Chen H., Cheuk R.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Kim C., Koseama E., Lam B., Lin J., Meyers M.C., Miranda M.,
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA Shinn P., Southwick A., Tracy S.E., Shinozaki K., Davis R.W.,
RA Ecker J.R., Theologis A.;
RT "Full Length cDNA of gene C7A10.390 (GI:4006876).";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY045987; AAK76561.1; -
DR InterPro; IPR005516; Remorin_C.
DR Pfam; PF03763; Remorin_C; 1.
KW Hypothetical protein.
SQ SEQUENCE 427 AA; 46788 MW; C8B0192D315DAE41 CRC64;

Query Match 54.3%; Score 44; DB 10; Length 427;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 ECKTPLTTAAAPVV 16
||| :||| :|||
Db 208 EKAEMTTAQSPVV 222
||| :||| :|||

RESULT 9
O23188 PRELIMINARY; PRT; 427 AA.
AC O23188;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN C7A10.390 OR AT4G36970.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1] :||| :||| :|||
RP SEQUENCE FROM N.A.
RA Bevan M., Terry N., Vos P., Heijnen L., Mewes H.W., Schueller C.,
RA Chaiwatwis N.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
[2]
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
RA Quach H.L., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
RA Yu G., Yuan S., Chen H., Cheuk R., Jones T., Kim C.J., Nguyen M.,
RA Palm C.J., Shinn P., Southwick A., Tripp M.G., Wu T., Davis R.W.,
RA Ecker J.R., Theologis A.;
RT "Arabidopsis Open Reading Frame (ORF) Clones.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z99707; CAB16794.1; -
DR EMBL: AL161590; CAB80363.1; -
DR EMBL: AY150393; AAN12938.1; -
DR InterPro; IPR005516; Remorin_C.
DR Pfam; PF03763; Remorin_C; 1.
KW Hypothetical protein.
SQ SEQUENCE 427 AA; 46800 MW; 75ACC8E02D90AE54 CRC64;

Query Match 54.3%; Score 44; DB 10; Length 427;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 ECKTPLTTAAAPVV 16
||| :||| :|||
Db 208 EKAEMTTAQSPVV 222
||| :||| :|||

RESULT 10
Q8KKH4 PRELIMINARY; PRT; 486 AA.
AC Q8KKH4;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Nitrite reductase precursor.
GN NIRK.
OS Hyphomicrobium denitrificans.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Hyphomicrobiaceae; Hyphomicrobium.
OX NCBI_TaxID=53399;
RN [1] :||| :||| :|||
RP SEQUENCE FROM N.A.
RC STRAIN=A3151;
RA Katoka K., Fukui A., Kabayashi M., Yamaguchi K., Suzuki S.;
RT "Cloning and expression of copper-containing nitrite reductase from
RL Hyphomicrobium denitrificans.";
RL EMBL: AB076606; BAC00912.1; -
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR001287; CuNO2 reductase.
DR PRINTS; PR00695; CUNO2RDTASE.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
KW Signal.
FT SIGNAL 1 39
SQ SEQUENCE 486 AA; 52238 MW; 6984E02ECE791AF CRC64;

Query Match 53.1%; Score 43; DB 2; Length 486;
Best Local Similarity 66.7%; Pred. No. 45;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ECKTPLTTAAAPVV 16
||| :||| :|||
Db 53 ECKTPTAGAHAPVV 67
||| :||| :|||

RESULT 11
P77924 PRELIMINARY; PRT; 506 AA.
AC P77924;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Catalase isozyme.
GN KATA.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=294;
RN [1] :||| :||| :|||
RP SEQUENCE FROM N.A.
RC STRAIN=Cb36;
RX MEDLINE=21318967; PubMed=11425481;
RA Peters M., Heinaru A., Nurf A.;
RT "plasmid-encoded catalase KATA, the main catalase of Pseudomonas
RL fluorescens strain Cb36.";
RL FEMS Microbiol. Lett. 200:235-240(2001).
DR EMBL; U72068; AAB17009.1; -
DR HSP; P42321; 1M85.
DR InterPro; IPR002226; Catalase.
DR Pfam; PF00199; catalase; 1.
DR PRINTS; PR00067; CATALASE.
DR ProDom; PD000510; Catalase; 1.
DR PROSITE; PS00438; CATALASE_2; 1.
KW Plasmid.
SQ SEQUENCE 506 AA; 57324 MW; F1EAA728C5D41CBE CRC64;
```

Query Match 53.1%; Score 43; DB 2; Length 506;
 Best Local Similarity 71.4%; Pred. No. 47;
 Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 TPLTTAAAPVVDN 18
 DB 11 TPLTTAAAPVVDN 24

RESULT 12

Q9DWH8 PRELIMINARY; PRT; 1240 AA.
 AC Q9DWH8; DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
 DE Pr2.
 GN R2.
 OS Rat cytomegalovirus (strain Maastricht).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Muromegalovirus.
 NCBI_TaxID=79700;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Maastricht;
 RX MEDLINE=20366325; PubMed=10906222;
 RA Vink C., Beuken E., Bruggeman C.A.;
 RT "Complete DNA sequence of the rat cytomegalovirus genome.";
 RL J. Virol. 74:7656-7665(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Maastricht;
 RX MEDLINE=20473137; PubMed=11018281;
 RA Gruijthuisen Y.K., Beuken E., Bruggeman C.A.;
 RT "rat cytomegalovirus R89 is a highly conserved gene which expresses a
 spliced transcript.";
 RL Virus Res. 69:119-130(2000).
 DR EMBL; AF232689; AAF9111.1;
 SQ SEQUENCE 1240 AA; 125612 MW; 33B6C13DC6A272B0 CRC64;

Query Match 53.1%; Score 43; DB 12; Length 1240;
 Best Local Similarity 64.3%; Pred. No. 1.2e+02;
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAPV 15
 DB 570 EERELTTAGAPV 583

RESULT 13

Q9RNU2 PRELIMINARY; PRT; 210 AA.
 AC Q9RNU2; DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
 DE Serum opacity factor precursor (Fragment).
 GN SOF4470.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=4470-96; TISSUE=Blood;
 RA Beall B., Gherardi G., Lovgren M., Tyrrell G., Facklam R., Forwick B.;
 RT "Predictions of M serotype, anti-opacity factor type, and highly
 related strain sets based upon the variable emm and sof gene
 sequences.";
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF179217; AAD55775.1;
 KW Signal.

FT NON_TER 1 1
 FT SIGNAL <1 7
 FT CHAIN 8 >210
 FT NON_TER 210 210
 SQ SEQUENCE 210 AA; 21220 MW; 199C89EECF260B6A CRC64;

Query Match 51.9%; Score 42; DB 2; Length 210;
 Best Local Similarity 57.1%; Pred. No. 29;
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAPV 15
 DB 103 EKTPTVTSTPV 116

RESULT 14

Q9X6X8 PRELIMINARY; PRT; 211 AA.
 AC Q9X6X8; DT 01-NOV-1999 (TReMBLrel. 12, Created)
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
 DE Serum opacity factor precursor (Fragment).
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SS1401 type PT1658;
 RA Beall B., Gherardi G.;
 RT "The relation of Streptococcus pyogenes sof and emm gene sequence
 RT types to genetically distinct strain sets.";
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF154330; AAD36988.1;
 KW Signal.

FT NON_TER 1 1
 FT SIGNAL <1 8
 FT CHAIN 9 >211
 FT NON_TER 211 211
 SQ SEQUENCE 211 AA; 21367 MW; F5475DDC6A084FE6 CRC64;

Query Match 51.9%; Score 42; DB 2; Length 211;
 Best Local Similarity 61.5%; Pred. No. 29;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAPV 14
 DB 98 EKTPTATSSAP 110

RESULT 15

Q9FT45 PRELIMINARY; PRT; 436 AA.
 AC Q9FT45; DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Hypothetical protein (GPI-anchored protein).
 GN T25B15.140.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Alcaraz J.P., Clabault G., Cottet A., Mache R., Mewes H.W., Rudd S.;
 RA Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;

RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.

RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=sv. Columbia;
 RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
 RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
 RA Hayashizaki Y., Shinozaki K.
 RT "Arabidopsis thaliana full-length cDNA."
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL132972; CAC07928.1; -.
 DR EMBL; AK117608; BAC42264.1; -.
 DR InterPro; IPR000782; BIGH3_fasciclin.
 DR Pfam; PF02469; Fasciclin; 1.
 DR SMART; SM00554; FAS1; 2.
 DR PROSITE; PS00976; NMT 2; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 436 AA; 48073 MW; 83FD94879F9855DF CRC64;

Query Match 51.9%; Score 42; DB 10; Length 436;
 Best Local Similarity 50.0%; Pred. No. 60;
 Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAPVVKXN 19
 |||||
 Db 395 EKTPEVKEKTVGVVKKA 412

RESULT 16

Q9RJK9 ID Q9RJK9 PRELIMINARY; PRT; 487 AA.
 AC Q9RJK9;
 DT 01-MAY-2000 (TREMELrel. 13, Created)
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
 DT 01-MAY-2003 (TREMELrel. 23, Last annotation update)
 DE Catalase (EC 1.11.1.6).
 GN KATA OR SC00379 OR SCF62.05.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Murphy L., Harris D.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;

RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 CC -!- FUNCTION: OCCURS IN ALMOST ALL AEROBICALLY RESPIRING ORGANISMS AND
 SERVES TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN
 PEROXIDE (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.

DR EMBL; AL939105; CAB58320.1; -.
 DR HSSP; P42321; 1M85.
 DR InterPro; IPR002226; Catalase.
 DR Pfam; PF00199; catalase; 1.
 DR PRINTS; PR00067; CATALASE.
 DR ProDom; PD000510; Catalase; 1.
 DR PROSITE; PS00437; CATALASE_1; 1.
 DR PROSITE; PS00438; CATALASE_2; 1.
 KW Heme; Hydrogen peroxide; Iron; Oxidoreductase; Peroxidase;
 KW Complete proteome.
 SQ SEQUENCE 487 AA; 55116 MW; 9D333489EAF60B7 CRC64;

Query Match 51.9%; Score 42; DB 16; Length 487;
 Best Local Similarity 69.2%; Pred. No. 67;
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 6 PLTTAAXAPVVKXN 18
 |||||
 Db 8 PLTTVAGAPVVDN 20

RESULT 17

Q9SE04 ID Q9SE04 PRELIMINARY; PRT; 132 AA.
 AC Q9SE04;
 DT 01-MAY-2000 (TREMELrel. 13, Created)
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
 DT 01-MAY-2003 (TREMELrel. 23, Last annotation update)
 DE Copper chaperone homolog CCH.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98369102; PubMed=9701579;
 RA Himelblau E., Mira H., Lin S.J., Culotta V.C., Penarrubia L.,
 RA Amasino R.M.;
 RT "Identification of a functional homolog of the yeast copper
 homeostasis gene ATX1 from Arabidopsis."
 RL Plant Physiol. 117:1227-1234(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Mira H., Penarrubia L.;
 RT "Copper chaperone from Oryza sativa."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF198626; AAF15285.1; -.
 DR HSSP; O00244; 1FE0.
 DR Gramene; Q9SE04;
 DR InterPro; IPR006121; HeavyMe transp.
 DR InterPro; IPR006191; Metal_bind.
 DR Pfam; PF00403; HMA; 1.
 DR PROSITE; PS0846; HMA 2; 1.
 SQ SEQUENCE 132 AA; 13094 MW; 7176EF95350A8231 CRC64;

Query Match 50.6%; Score 41; DB 10; Length 132;
 Best Local Similarity 50.0%; Pred. No. 26;
 Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 EKTPLTTAAXAPVVKXN 19
 |||||
 Db 97 EAAPPTTAAEAPATAA 114

RESULT 18

```
Q54272
ID Q54272 PRELIMINARY; PRT; 417 AA.
AC Q54272;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Streptomyces hygroscopicus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1912;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=SF1293;
RX MEDLINE=95309717; PubMed=7789803;
RA Hidaka T., Hidaka M., Kuzuyama T., Seto H.;
RT "Sequence of a P-methyltransferase encoding gene isolated from a
RT bialaphos-producing streptomycetes hygroscopicus.";
RL Gene 158:149-150(1995);
CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR EMBL; D37877; BAA07116.1; -;
DR HSSP; P56533; 1A4S.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Hypothetical protein; Oxidoreductase.
SQ SEQUENCE 417 AA; 43406 MW; 9E8714F4FB654B47 CRC64;

Query Match 50.6%; Score 41; DB 2; Length 417;
Best Local Similarity 52.9%; Pred. No. 85;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 EKTPLTTAAXAPVXNA 19
Db 117 EKTPLTALAFALLIEA 133

RESULT 19
Q8P7W8 PRELIMINARY; PRT; 516 AA.
AC Q8P7W8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE PilY1 protein.
GN PILY1 OR XCC2489.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Camanav A.F., Cardoso J., Chamberg F., Chapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Borri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.P., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Nachado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira B.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
EMBL; AE012360; AAM41763.1; -;

KW Complete proteome.
SQ SEQUENCE 516 AA; 53416 MW; 88A17A8B40B96578 CRC64;

Query Match 50.6%; Score 41; DB 16; Length 516;
Best Local Similarity 72.7%; Pred. No. 1.1e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 TPLTTAAXAPV 15
Db 454 TPLTTSGNAPV 464

RESULT 20
Q8ZZX0 PRELIMINARY; PRT; 663 AA.
AC Q8ZZX0;
DT 01-WAR-2002 (TrEMBLrel. 20, Created)
DT 01-WAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Conserved within P. aerophilum.
GN PAE0040.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT aerophilum";
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR EMBL; AE009747; AAL62519.1; -;
DR InterPro; IPR001064; Crystallin.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
KW Complete proteome.
SQ SEQUENCE 663 AA; 71372 MW; 216F07B821B78B76 CRC64;

Query Match 50.6%; Score 41; DB 17; Length 663;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 9; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 BEKPLTTAAXAPVXNA 19
Db 617 ESATPATTATATPKEDNA 634

RESULT 21
Q8WT63 PRELIMINARY; PRT; 920 AA.
AC Q8WT63;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE RE03056P.
GN CG30480 OR CG18373.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscophora;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RT Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
```

```
DR EMBL; AY118357; AAM48386.1; --
DR FlyBase; FBgn0050480; CG30480.
DR InterPro; IPR002052; N6_Mtase.
DR PROSITE; PS00092; N6_MTASE; 1.
SQ SEQUENCE 920 AA; 101042 MW; 34FF1CBB293AF79 CRC64;

Query Match 50.6%; Score 41; DB 5; Length 920;
Best Local Similarity 53.3%; Pred. No. 1.9e+02;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 EKTPTLTTAAXAPVV 16
Db 695 EEVTPYNTAPHSPIV 709

RESULT 22
Q8ML93 PRELIMINARY; PRT; 920 AA.
AC Q8ML93;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG30480-PA.
GN CG30480 OR CG18373.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karp G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sidenkian I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapieton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
```



```

DR EMBL; 282273; CAB54978.1; -.
DR WormPep; F56F12.1; CR23785.
DR InterPro; IPR003127; Sorb.
DR SMART; SM00459; Sorb; 1.
SQ SEQUENCE 171 AA; 18844 MW; D782482A786F7941 CRC64;

Query Match 49.4%; Score 40; DB 5; Length 171;
Best Local Similarity 53.3%; Pred. No. 51;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 5 TPLTTAAAXAPVVXNA 19
DB 73 TPTTTAAPIAAVA 87

RESULT 24
Q9U4Y3
ID Q9U4Y3 PRELIMINARY; PRT; 212 AA.
AC Q9U4Y3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Fruitless protein (Fragment).
GN FRU.
OS Ceratitis capitata (Mediterranean fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Tephritoidea; Tephritidae; Ceratitis.
OX NCBI_TaxID=7213;
RN [1]
RP SEQUENCE FROM N.A.
RA Davis T., Yamamoto D.;
RT "Sequence of the fruitless gene in Ceratitis capitata.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF124047; RAP22527.1; -.
DR InterPro; IPR00210; BTB_POZ.
DR Pfam; PF00651; BTB; 1.
DR SMART; SM00225; BTB; 1.
DR PROSITE; PS00097; BTB; 1.
FT NON_TER 1
FT NON_TER 212
FT NON_TER 212
SQ SEQUENCE 212 AA; 22715 MW; D1B7C39FBC5A8AB CRC64;

Query Match 49.4%; Score 40; DB 5; Length 212;
Best Local Similarity 52.9%; Pred. No. 63;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 EKTPLTTAAAXAPVVXNA 19
DB 188 ERSPLTAAAMAAVA 204

RESULT 25
Q8PQ46
ID Q8PQ46 PRELIMINARY; PRT; 238 AA.
AC Q8PQ46;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein XAC0482.
GN XAC0482.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=2022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL; AE012145; AAM39789.1; -.
DR InterPro; IPR005834; Hydrolase.
DR Pfam; PF00702; Hydrolase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 238 AA; 26450 MW; B9D14734696D0568 CRC64;

Query Match 49.4%; Score 40; DB 16; Length 238;
Best Local Similarity 69.2%; Pred. No. 71;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL; AE012145; AAM39789.1; -.
DR InterPro; IPR005834; Hydrolase.
DR Pfam; PF00702; Hydrolase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 238 AA; 26450 MW; B9D14734696D0568 CRC64;

Query Match 49.4%; Score 40; DB 16; Length 238;
Best Local Similarity 69.2%; Pred. No. 71;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 6 PLTTAAAXAPVVXN 18
DB 194 PLLAAARAPVVVN 206

RESULT 26
Q8PD69
ID Q8PD69 PRELIMINARY; PRT; 238 AA.
AC Q8PD69;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein XCC0471.
GN XCC0471.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=2022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL; AE012145; AAM39789.1; -.
DR InterPro; IPR005834; Hydrolase.
DR Pfam; PF00702; Hydrolase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 238 AA; 26450 MW; B9D14734696D0568 CRC64;

Query Match 49.4%; Score 40; DB 16; Length 238;
Best Local Similarity 69.2%; Pred. No. 71;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 6 PLTTAAAXAPVVXN 18
DB 194 PLLAAARAPVVVN 206

```

```

QY 6 PLTTAAXAPVVXN 18
DB 194 PLTKAARAPVVVN 206

RESULT 27
QY9984 PRELIMINARY; PRT; 252 AA.
ID QY9984
AC QY9984;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical protein APE2401.
GN APE2401.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000064; BAA81416.1; -.
SQ SEQUENCE 252 AA; 25912 MW; 527C1D8A70FFDC14 CRC64;

Query Match 49.4%; Score 40; DB 17; Length 252;
Best Local Similarity 88.9%; Pred. No. 75;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 PLTTAAXAP 14
DB 64 PLTTAASAP 72

RESULT 28
QY5M9 PRELIMINARY; PRT; 258 AA.
AC QY5M9;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein lmo2028.
GN LMO2028.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Ruiniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Duesurget O.,
RA Etienne K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Gobel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kerst U., Kretz J., Kuhn M., Kunst F., Kurapkut G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";

RL Science 294:849-852(2001).
DR EMBL; AL591982; CAD00106.1; -.
DR Listlist; LMO2028; -.
DR InterPro; IPR002942; S4.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 258 AA; 29682 MW; AB62C55270CFD90F CRC64;

Query Match 49.4%; Score 40; DB 16; Length 258;
Best Local Similarity 57.1%; Pred. No. 77;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 EKTPLTTAAXAPVV 16
DB 158 EETELTNAVYAPVI 171

RESULT 29
Q811A1 PRELIMINARY; PRT; 367 AA.
ID Q811A1
AC Q811A1;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE EVE-PA.
GN EVE.
OS Drosophila littoralis.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=47316;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tucson 15010-1001.10;
RX MEDLINE=22426072; PubMed=12537575;
RA Bergman C.M., Pfeiffer B.D., Rincon-Limas D.E., Hoskins R.A.,
RA Gnirke A., Mungall C.J., Wang A.M., Kronmiller B., Pacleb J., Park S.,
RA Stapleton M., Wan K., George R.A., de Jong P.J., Botas J., Rubin G.M.,
RA Celniker S.E.;
RT "Assessing the impact of comparative genomic sequence data on the
RT functional annotation of the drosophila genome.";
RL Genome Biol. 3:RESEARCH0086.1-RESEARCH0086.20(2002).
DR EMBL; AY190953; AA01071.1; -.
SQ SEQUENCE 367 AA; 39464 MW; FDC6800F7040121F CRC64;

Query Match 49.4%; Score 40; DB 5; Length 367;
Best Local Similarity 53.3%; Pred. No. 1.1e+02;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 TELTTAAXAPVVXNA 19
DB 303 SPLRSAPTAPAVANA 317

RESULT 30
Q9A4B0 PRELIMINARY; PRT; 394 AA.
ID Q9A4B0
AC Q9A4B0;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Hypothetical protein CC2927.
GN CC2927.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,

```

RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utlarback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,
RT "Complete genome sequence of *Caulobacter crescentus*."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141 (2001).
DR TIGR; CC2927; -
DR InterPro; IPR005625; DUF337.
DR Pfam; PF03929; DUF337; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 394 AA; 42675 MW; 8B1178E174DD18D CRC64;

Query Match 49.4%; Score 40; DB 16; Length 394;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 4 KTPLTTAAAXAPVVXNA 19
||||| : : :
Db 315 KTPLTTTHLTPALIDA 330

RESULT 31
Q9RD97 PRELIMINARY; PRT; 477 AA.
AC Q9RD97;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative transmembrane efflux protein.
GN SC02854 OR SCE20.28.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96 (1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdono-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)."
RL Nature 417:141-147 (2002).
DR EMBL; AL939114; CAB65584.1; -.

DR InterPro; IPR007114; MFS.
DR InterPro; IPR005829; Sug. transporter.
DR PROSITE; PS50850; MFS; 1.
DR PROSITE; PS00216; SUGAR TRANSPORT_1; 1.
KW Transmembrane; Complete proteome.
SQ SEQUENCE 477 AA; 47686 MW; 20797B126B148090 CRC64;

Query Match 49.4%; Score 40; DB 16; Length 477;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 TPLTTAAAXAP 14
||||| : : :
Db 19 TPATTAAXAP 28

RESULT 32
Q99665 PRELIMINARY; PRT; 498 AA.
ID Q99665
AC Q99665;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Gag polyprotein.
GN GAG.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94149848; PubMed=8107219;
RA Gurtler L.G., Hauser P.H., Eberle J., von Brunn A., Knapp S.,
RA Zekeng L., Tsague J.M., Kaptue L.;
RT "A new subtype of human immunodeficiency virus type 1 (MVP-5180) from
Cameroon."
RL J. Virol. 68:1581-1585 (1994).
DR EMBL; L20571; AAA44859.1; -.
DR HSP; P05888; 1AAF.
DR InterPro; IPR000721; Gag p24.
DR InterPro; IPR000071; Retrovir_p17.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00540; Gag_p17; 1.
DR Pfam; PF00607; Gag_p24; 1.
DR Pfam; PF00098; zf_CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR PRINTS; PR00234; HIV1MATRIX.
DR SMART; SM00343; Znf_C2HC; 2.
DR PROSITE; PS50158; ZF_CCHC; 2.
KW AIDS; Core protein; Polyprotein.
SQ SEQUENCE 498 AA; 55498 MW; 6DE3A29B4E6A17D4 CRC64;

Query Match 49.4%; Score 40; DB 15; Length 498;
Best Local Similarity 44.4%; Pred. No. 1.5e+02;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 EEKPLTTAAAXAPVVXNA 19
||||| : : :
Db 119 EETSPTQTSQNYPIVNTA 136

RESULT 33
Q9BLH5 PRELIMINARY; PRT; 594 AA.
ID Q9BLH5
AC Q9BLH5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE HRETR-1 protein.
GN HRETR-1.
OS Halocynthia roretzi (Sea squirt).
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
OC Stolidobranchia; Pyuridae; Halocynthia.
OX NCBI_TaxID=7729;

```

RN SEQUENCE FROM N.A.
RP Yagi K., Makabe K.W.;
RA "New neural marker genes expressing in subsets of embryonic neural
RT cells of the ascidian, Halocynthia roretzi.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047035; BAB40781.1; -
DR HSP; FL9339; 2SXL.
DR InterPro; IPR002343; Hud_Sx1_RNA.
DR InterPro; IPR00504; RNA_rec_mot.
DR Pfam; PF00076; rtm; 3.
DR PRINTS; PR00961; HUDSKLRNA.
DR SMART; SM00360; RM; 3.
DR PROSITE; PS50302; RM; 3.
SQ SEQUENCE 594 AA; 63686 MW; 193CDE2B14216920 CRC64;

Query Match          49.4%; Score 40; DB 5; Length 594;
Best Local Similarity 46.7%; Pred. No. 1.8e+02;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 TPLTTAAAXAPVXNA 19
DB 483 TPTTTTQCAPLVNS 497

RESULT 34
Q9W180 ID Q9W180 PRELIMINARY; PRT; 788 AA.
AC Q9W180;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-WAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE MACR-60C protein.
GN MACR-60C OR CG4356.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
RA de Fabios B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

```

```

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL; AB003464; AAF47197.1; -.
DR FlyBase; FBgn0000037; mACR-60C.
DR InterPro; IPR00276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHOOPS.
DR PROSITE; PS00237; G-PROTEIN_RECEP_F1_1; 1.
DR PROSITE; PS0262; G-PROTEIN_RECEP_F1_2; 1.
DR PROSITE; PS50362; G-PROTEIN_RECEP_Receptor; Transmembrane.
KW G-protein coupled receptor; Receptor; Transmembrane.
SQ SEQUENCE 788 AA; 84791 MW; AD8476D9F52AB522 CRC64;

Query Match          49.4%; Score 40; DB 5; Length 788;
Best Local Similarity 52.9%; Pred. No. 2.4e+02;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 EKTPLTTAAAXAPVXNA 19
DB 602 EDGPTTTAAAPLASAA 618

RESULT 35
Q8MLP2 ID Q8MLP2 PRELIMINARY; PRT; 805 AA.
AC Q8MLP2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG4356-PB.
GN MACR-60C OR CG4356.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA de Fabios B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

```

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith H.,
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong J., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celnikier S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Anantides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pachet J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of *Drosophila melanogaster* genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celnikier S.E.,
RA Clamp M., Drysdale R., Emert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of *Drosophila melanogaster* genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celnikier S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003464; AAM68310.1; -;
DR FlyBase; FBgn0000037; mACR-60C.
DR InterPro; IPR000276; GPCR_Rhodopsin.
DR Pfam; PF00001; 7tm 1; 1.
DR PRINTS; PRO0237; GPCRHOOPS.
DR PROSITE; PS00237; G PROTEIN RECEPTOR F1.1; 1.
DR PROSITE; PS0262; G PROTEIN RECEPTOR F1.2; 1.
SQ SEQUENCE 805 AA; 86623 MW; 97A329CAA5BBED8 CRC64;

Query Match 49.4%; Score 40; DB 5; Length 805;
Best Local Similarity 52.9%; Pred. No. 2.5e+02;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 EKTPLTTAAAPVXNA 19
| | | | | | | | | |
Db 619 EDGPTTAAAPLASAA 635

RESULT 36
O57577 PRELIMINARY; PRT; 846 AA.
AC O57577;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE NCAM-140.
OS Cynops pyrrhogaster (Japanese common newt).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.
OX NCBI_TaxID=8330;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Mizuno M., Takabatake T., Takahashi T.C., Takeshima K.;
RT "pax-6 gene expression in newt eye development.";
RL Dev. Genes Evol. 207:167-176 (1997).
DR EMBL; D85084; BAA24033.1; -;
DR HSSP; P56276; ITLK.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG C2.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00047; ig; 5.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00408; IGC2; 5.
DR PROSITE; PS50835; IG LIKE; 3.
KW Immunoglobulin domain.
SQ SEQUENCE 846 AA; 92544 MW; 59D599C035A1DDB3 CRC64;

Query Match 49.4%; Score 40; DB 13; Length 846;
Best Local Similarity 52.9%; Pred. No. 2.6e+02;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 EKTPLTTAAAPVXNA 19
| | | | | | | | | |
Db 799 ETTPLTEPEKAPVEENS 815

RESULT 37
Q8CGF1 PRELIMINARY; PRT; 1266 AA.
ID Q8CGF1;
AC Q8CGF1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to PTP1L-associated RhoGAP 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N;
RA Strausberg R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC040387; RAH40387.1; -;
SQ SEQUENCE 1266 AA; 142340 MW; 96687B52C3D229D2 CRC64;

Query Match 49.4%; Score 40; DB 11; Length 1266;
Best Local Similarity 50.0%; Pred. No. 3.9e+02;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAPVXNA 19
| | | | | | | | | |
Db 1185 EKTALPSIAVPPVLVHA 1202

RESULT 38
O02006 PRELIMINARY; PRT; 1360 AA.
ID O02006;
AC O02006;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Fruit fly retrotransposon ninja DNA, complete sequence (Fragment).
DE NINJA\POL.

```

OS Drosophila simulans (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7240;
RX STRAIN=white;
RP SEQUENCE FROM N.A.
RC STRAIN=white-chocolate; TISSUE=Body;
RX MEDLINE=96252483; PubMed=8925473;
Ogura K., Takeuchi S., Nakayama T., Yamamoto M.;
RT "Molecular structure of the transposable element ninja in Drosophila
RT simulans.";
RL Genes Genet. Syst. 71:1-8(1996).
DR EMBL; D83207; BAA57030.1; -
DR MEROPS; A17.001; -
DR FlyBase; FBgn0044186; Dsim\ninja\pol.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf-CCHC; 1.
DR SMART; SM00343; Znf_C2HC; 1.
DR PROSITE; PS00158; ZF_CCHC; 1.
FT NON_TER 1
SQ SEQUENCE 1360 AA; 150936 MW; DF2021C8DA4FAE0 CRC64;

Query Match 49.4%; Score 40; DB 5; Length 1360;
Best Local Similarity 47.1%; Pred. No. 4.2e+02;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 EKTPLTTAAAXAPVXNA 19
Db - |||:|:|:|
101 ENTPIITTAQQPASSGA 117

RESULT 39
Q9VHP9 PRELIMINARY; PRT; 1399 AA.
AC Q9VHP9; Q9VHP8;
DT 01-WAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG31258 protein.
GN CG31460 OR CG11745.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flossler A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.F., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

```

```

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J., Wang X.,
RA Williams S.M., Woodedge T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celnik S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.B., Doyle C., Drenek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragov V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celnik S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise B., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celnik S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003679; AAF54252.3; -
DR FlyBase; FBgn0051460; CG31460.
DR InterPro; IPR004927; TF_bZIP.
DR PROSITE; PS00036; BZIP_BASIC; 1.
SQ SEQUENCE 1399 AA; 156378 MW; B18E1A86714A895E CRC64;

Query Match 49.4%; Score 40; DB 5; Length 1399;
Best Local Similarity 53.3%; Pred. No. 4.3e+02;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 EKTPLTTAAAXAPV 16
Db - |||:|:|:|
1241 DEQAPYTEAALGPV 1255

RESULT 40
Q03462 PRELIMINARY; PRT; 405 AA.
AC Q03462;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

```

DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE OPAQUE2 HETERODIMERIZING protein 1 (OHP1).
GN OHP1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R802; TISSUE=ENDOSPERM;
RX MEDLINE=93200807; PubMed=8453304;
RA Fyish L.D., Aukerman M.J., Schmidt R.J.;
RT "OHP1: a maize basic domain/leucine zipper protein that interacts with
RT opaque2.";
RL Plant Cell 5:227-236(1993).
CC -!- FUNCTION: A TRANS-ACTING TRANSCRIPTIONAL ACTIVATOR OF 22 KDA ZEIN
CC GENES. BINDS TO THE O2 SITE, CONSENSUS SEQUENCE: GATGAYRTGR.
CC -!- SUBUNIT: BINDS DNA AS A HOMODIMER OR AS A HETERODIMER WITH
CC OPAQUE2.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: TO OTHER BZIP PROTEINS.
DR EMBL; L00623; AAA33488.1; --
DR TRANSFAC; T02721; --
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF00170; bZIP; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00036; BZIP_BASIC; 1.
KW Transcription regulation; Activator; DNA-binding; Nuclear protein;
KW Trans-acting factor.
FT DOMAIN 240 282 LEUCINE-ZIPPER.
FT DNA_BIND 214 233 BASIC MOTIF.
SQ SEQUENCE 405 AA; 42320 MW; 3983CAAEBEC90D1 CRC64;

Query Match 48.8%; Score 39.5; DB 10; Length 405;
Best Local Similarity 48.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 6; Indels 7; Gaps 1;

OY 2 EKTPLTTAAXA-----PVVKNA 19
Db 79 EKTPTCTAAAAAASSSVDPVEYNA 103

Search completed: January 28, 2004, 13:09:22
Job time : 36 secs

THIS PAGE BLANK (uspto)